KUHLMANN, BEATRICE G., Ph.D. To Use or not to Use? Influences of List Presentation Format and Working Memory Capacity on Older Adults' Semantic Clustering (2013)
Directed by Dr. Dayna R. Touron. 121 pp.

The goal of the present study was to examine effects of list presentation format (study list presented as a whole vs. words presented briefly individually) on younger and older adults' semantic clustering of study words. Spontaneous clustering use did not differ between format conditions in either age group. Older adults spontaneously clustered to a similar extent as younger adults, evidencing no production deficiency. When clustering use was instructed, the whole-list format conditions clustered more successfully, resulting in greater recall than in the individual-words conditions, even under dual-task demands. Older adults clustered less successfully than younger adults, evidencing a utilization deficiency, with no overall recall improvements after clustering instructions in the individual-words format. Clustering interfered with performance on a simple tone-discrimination task, indicating its general cognitive resource demands; absolute interference was greater for older adults. Working memory capacity (WMC) predicted clustering success and mediated age-related reductions therein when clustering use was instructed but not for spontaneous use. WMC-clustering correlations were similar across presentation formats but adjusted means differed such that individuals at the same level of WMC clustered more successfully in the whole-list format. Beliefs about clustering difficulty correlated with its spontaneous use but did not evidence metacognitive awareness of presentation-format effects in either age group. These results suggest that a simple change in presentation format can facilitate encoding strategy use,

particularly for older adults, but these benefits do not necessarily translate into spontaneous use differences. Thereby, presentation format alone cannot explain mixed findings regarding age-related differences in spontaneous clustering.

TO USE OR NOT TO USE? INFLUENCES OF LIST PRESENTATION FORMAT AND WORKING MEMORY CAPACITY ON OLDER ADULTS' SEMANTIC CLUSTERING

by

Beatrice G. Kuhlmann

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2013

Approved by
Dayna R. Touron______________
Committee Chair

APPROVAL PAGE

This dissertation written by Beatrice G. Kuhlmann has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ____
Dayna R. Touron

Committee Members ____
Robert E. Guttentag

____
Michael J. Kane

____
Stuart Marcovitch

2/27/2013____________________
Date of Acceptance by Committee

2/27/2013_________________
Date of Final Oral Examination

ACKNOWLEDGMENTS

I thank my advisor, Dr. Dayna R. Touron, for her mentorship throughout the past years. This dissertation project was inspired by her thinking about age-related differences in cognitive performance. I also thank my former and current committee members, Drs. Robert E. Guttentag, Michael J. Kane, Stuart Marcovitch, and Lili Sahakyan for sharing their expertise and helping me sharpen my research ideas. Further I thank Dr. Douglas W. Levine for advising me on some of the statistical analyses and helping me think about the clustering measures.

I received funding for this project through a Dissertation Research Award from the American Psychological Association (APA). I appreciate the help of Christopher Dillon, Lauren Harden, and Ashley Jennings with participant recruitment and data collection. Special thanks to the older adult participants for coming to campus and volunteering their time.

I was fortunate to spend graduate school with a supportive group of fellow graduate students. I have greatly benefitted from discussions and the exchange of ideas with the other “coggies.” Finally, I thank my family and friends, near and far, for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Declines in cognitive performance, including episodic memory, are inevitable with increasing adult age, even if one is spared from dementia (see Park, 2000, for a review). Nonetheless, cognitive performance remains plastic in old age (Baltes & Baltes, 1993; Hertzog, Kramer, Wilson, & Lindenberger, 2008). For example, in a meta-analysis by Verhaeghen, Marcoen, and Gossens (1992), mean memory improvements in older adults through the use of encoding strategies were large (i.e., 0.78 standard deviations). Effective encoding strategies improve memory by fostering the deep, meaning-based elaboration and interrelation of to-be-studied material (Craik & Lockhart, 1972; Hunt & Einstein, 1981), such as semantic clustering of study words or the generation of mental images or sentence/stories (e.g., Kausler, 1994). Several studies found older adults to spontaneously use such effective encoding strategies less frequently than younger adults, both on memory tasks in the laboratory (e.g., Dunlosky & Hertzog, 2001; Taconnat et al., 2009) and in daily life (Bouazzaoui et al., 2010). Given the general benefit of encoding strategies in older adults, such a *production deficiency* exaggerates observed age-related memory differences which often become smaller when equivalent encoding strategy use among younger and older adults is ensured (e.g., Dunlosky & Hertzog, 2001; Hultsch, 1971; Naveh-Benjamin, Brav, & Levy, 2007). Some encoding strategies, however, are less beneficial in terms of memory improvement for older compared to younger adults. In the case of such a *utilization deficiency*, age-related memory differences become even larger when the strategy is used (e.g., Kliegl, Smith, & Baltes, 1989, 1990; Mason & Smith, 1977; Verhaeghen & Marcoen, 1996). That is, encoding strategy deficiencies can limit older adults' memory performance and contribute to age-related differences in memory. The goal of the present study was to better understand when and why such deficiencies occur by examining the influence of word-list presentation format on older adults' semantic clustering.

Variations in Older Adults' Strategic Abilities

Research on older adults' encoding strategy production and utilization has yielded mixed results, even regarding the same particular strategy. For example, several studies reporting age-related reductions in the use of semantic clustering (Amrhein, Bond, & Hamilton, 1999; Howard, McAndrews, & Lasaga, 1981; Hultsch, 1971; Jacobs, Rakitin, Zubin, Ventura, & Stern, 2001; Mungas, Ehlers, & Blunden, 1991; Schneider & Uhl, 1990; Taconnat et al., 2009; Wegesin, Jacobs, Zubin, Ventura, & Stern, 2000) are opposed by a similarly large group of studies finding that older adults use semantic clustering at least as much as younger adults (Blatt-Eisengart & Lachman, 2004; Hertzog, Dixon, & Hultsch, 1990; Hess, Auman, Colcombe, & Rahhal, 2003; Lachman & Andreoletti, 2006; Rankin, Karol, & Tuten, 1984; West, Dark-Freudeman, & Bagwell, 2009; Witte, Freund, & Brown-Whistler, 1993; Zivian & Darjes, 1983). Two studies only found lower clustering use in older adults under dual-task demands (Fernandes & Grady, 2008; Park, Smith, Dudley, & Lafronza, 1989) and one only when performance goals had been set (West & Thorn, 2001). But little is known about less obvious task factors that might moderate the occurrence of an age-related production deficiency for semantic clustering and thereby explain the mixed findings. Research on age-related differences in use of other encoding strategies has identified some task factors that act as moderators. Specifically, presentation format moderates age-related differences in self-testing (Bottiroli, Dunlosky, Guerini, Cavallini, & Hertzog, 2010) and word concreteness influences age-related differences in imagery and sentence use (Rowe & Schnore, 1971; Tournier & Postal, 2011).

Similarly, findings are mixed regarding older adults' memory benefits from particular encoding strategies. Even though semantic clustering use is generally positively correlated with recall in both younger and older adults (e.g., Hess et al., 2003; Lachman & Andreoletti, 2006), Taconnat et al. (2009) found no or even negative correlations in older adults. For a multi-step number-consonant mnemonic, older adults' memory for four-digit numbers only improved when they wrote down each step at encoding and retrieval (Derwinger, Stigsdotter Neely, & Bäckman, 2005). Similarly, Cavallini, Pagnin, & Vecchi (2003) found that older adults improved less than younger adults after training the method of loci or imagery-based strategies when tested on standard laboratory word lists but equally when tested with a more ecologically valid same-length list of grocery items. That is, task factors can optimize older adults' memory benefits from an encoding strategy.

Some researchers suggest that older adults' strategy use on cognitive tasks is particularly sensitive to the tasks' *affordability* of strategy use – that is how "easy" a task renders the use of a strategy (Bottiroli et al., 2010; Touron & Hertzog, 2004b). However, it remains unclear how a task makes strategy use more or less affordable and why older adults might be particularly sensitive to such manipulations. I propose that task factors influence the cognitive costs of encoding strategy execution such that certain task set-ups render the use of an encoding strategy less demanding on general cognitive resources. Given pronounced age-related declines in general cognitive resources like processing speed and working memory capacity (WMC; see Park, 2000 for a review), encoding strategies may be particularly taxing on older adults' available resources, making their use particularly effortful for the old. Therefore, strategy deficiencies in older adults might be caused by age-related declines in general cognitive resources and consequently older adults might particularly benefit from task factors that alleviate the resource demands of strategy use.

Cognitive Costs of Encoding Strategy Execution

There is good evidence that encoding strategies are indeed demanding of general cognitive resources and thereby particularly taxing on older adults. The most direct evidence stems from measuring the interference between strategic encoding and a simultaneously executed simple secondary task (Naveh-Benjamin, Craik, Guez, & Kreuger, 2005; see Guttentag, 1984, and Kee & Davies, 1990, 1991, for such evidence with children). Naveh-Benjamin et al. used such a dual-task interference paradigm to measure and compare cognitive costs of image and sentence generation for noun pairs in younger and older adults. In particular, they compared younger and older adults' performance on a simple visual tracking task when performed alone to when performed while strategically encoding. For both age groups, strategic encoding interfered more with tracking performance than did noun-pair encoding in comparison groups where no strategy use was instructed. That is, strategic encoding was more taxing on general cognitive resources also needed for performance on the tracking task than was non-strategic encoding. Further, this interference was stronger for older than for younger adults, suggesting that strategy execution was more taxing on older adults' available general cognitive resources.

Further evidence comes from correlations of strategy use with measure of general cognitive resources. Verhaeghen and Marcoen (1994) report positive correlations between basic processing speed and various encoding strategies, including semantic clustering, self-testing, repetition, image and sentence generation, in a sample of younger and older adults. They further showed that processing speed mediated the age-related reduction in production of these strategies. Additionally, Bouazzaoui et al. (2010) found that age-related differences in the reported frequency of using various encoding strategies in daily life were fully mediated by executive functioning. For semantic clustering use, positive correlations have been reported with executive functioning measures (in younger and older adults; Jacobs et al., 2001; Taconnat et al., 2009) as well as with WMC (in older adults only; Wegesin et al., 2000). Further evidence that semantic clustering is demanding of general cognitive resources comes from the finding that clustering use decreases under divided attention at encoding or retrieval (Fernandes & Grady, 2008; Park et al., 1989). Interestingly, Park et al. only found stronger decreases in clustering use in older compared to younger adults under dual task at encoding but not at retrieval, suggesting that semantic grouping of study words during encoding is particularly taxing for older adults.

For some strategies, older adults' reduced cognitive resources leave them unable to successfully execute the strategy, resulting in a utilization deficiency. For example, age-related declines in processing speed and WMC mediate the reduced memory benefit from the method of loci in older adults (Kliegl et al., 1990; Verhaeghen & Marcoen, 1996). Production deficiencies, however, occur even when older adults' resources are sufficient for succesful strategy execution, evidenced by age-equivalent memory benefits of the strategy when used (e.g., Dunlosky & Hertzog, 2001; Naveh-Benjamin et al., 2007). Indeed, in Naveh-Benjamin et al.'s (2005) study older adults experienced greater costs of encoding strategy use but memory improvements through strategies were comparable across age groups. In children, spontaneous use of imagery and sentence generation has been found to be negatively related to such a secondary-task cost-measure (Kee & Davies, 1990, 1991; see also Guttentag, 1984, for a similar finding with cumulative rehearsal). Naveh-Benjamin et al. (2005) did not assess spontaneous strategy use in addition to cognitive costs but it seems plausible that older adults' increased cognitive costs of encoding strategy execution cause them to use these strategies less frequently. Interestingly, it has been proposed that older adults might be particularly conservative in their expenditure of cognitive resources, resulting in the largest age-related differences when self-initiated effortful processing is involved in a task (Craik & Byrd, 1982; Hess, Germain, Swaim, & Osowski, 2009). Further, Brigham and Pressley (1988) found that older adults expressed cognitive effort concerns more frequently than younger adults when asked to choose between two encoding strategies. Consequently, there may be a particularly strong negative relationship between cognitive costs and strategy use in older adults.

The idea that cognitive costs influence strategy production is in line with formal models of strategy choice which posit that performance aspects of strategies influence which strategy an individual uses on a particular problem. In particular, strategy users are assumed to adaptively choose the best strategy for a task from their strategy repertoire in a cost-benefit analysis, considering each strategy's performance aspects including execution time, accuracy, and cognitive resource requirements (e.g., Lovett & Anderson, 1996; Shrager & Siegler, 1998; Siegler & Shipley, 1995). Children's and adults' strategy choices for arithmetic problems have been empirically demonstrated to be quite adaptive, with individuals choosing the strategy that is more accurate and faster for a given problem (Siegler, 1996; Siegler & Lemaire, 1997). The findings reviewed above suggest that production deficiencies in older adults might be related to their higher cognitive costs of encoding strategy execution but an empirical test of this question involving assessment of spontaneous strategy use alongside cognitive costs is needed. Importantly, a cognitive cost explanation of strategy use could also explain the influence of task factors on strategy production and utilization by considering how task factors might alter a strategy's resource demands.

Do Task Factors Alter Cognitive Costs of Strategy Execution?

Cognitive tasks greatly differ in their support of mental operations and age-related differences in cognitive performance are diminished the more supportive a task is; for example, a recall memory tasks requires more self-initiated processing than a recognition memory task and age-related memory difference are larger in the former than in the latter (Craik, 1986). Similarly, it has been reasoned that task factors may influence the cognitive demands of using a particular strategy and thereby moderate age-related differences in strategy use (Mata, von Helversen, & Rieskamp, 2010; Siegler & Lemaire, 1997; Touron & Hertzog, 2004b). With regards to encoding strategies, properties of the to-be-studied words have received the most research attention. Rowe and Schnore (1971) found that older adults' imagery production was particularly affected by the ease with which images can be generated for words (based on normative ratings; see also Hulicka & Grossman, 1967; Tournier & Postal, 2011). Witte et al. (1993) found age-equivalent increases in semantic clustering with more typical category exemplars (see also Howard et al., 1981).

A less recognized factor is presentation format of the study material (but see Bottiroli et al., 2010), which I propose to play an important role for older adults' production of semantic clustering. Usually, words from different semantic categories on a study list are presented randomly intermixed such that organizing by category is quite demanding on cognitive resources, requiring maintenance and reordering of already presented words while simultaneously evaluating their semantic relationships to newly presented words (cf., Jacobs et al., 2001; Taconnat et al., 2009). Jacobs et al. found that presenting words from the same category in succession (i.e., blocked), as opposed to randomly interspersed throughout the study list, significantly increased spontaneous use of semantic clustering in both younger and older adults. Across studies on aging and semantic clustering, a notable methodological variation in presentation format is that some studies briefly present the words (not blocked by category) individually (Amrhein et al., 1999; Fernandes & Grady, 2008; Jacobs et al., 2001; Mungas et al., 1991; Park et al., 1989; Taconnat et al., 2009; Wegesin et al., 2000; Witte et al., 1993) whereas others present the whole study list (again not blocked by category) at once for a longer duration (Blatt-Eisengart & Lachman, 2004; Hertzog et al., 1990; Hess et al., 2003; Lachman & Andreoletti, 2006; Rankin et al., 1984; Schneider & Uhl, 1990; West et al., 2009; West & Thorn, 2001). Most notably, it is primarily the studies using the former individual-words format that find age-related differences in spontaneous clustering use (with the exception of Witte et al., 1993, and only under dual task in Fernandes & Grady, 2008, and Park et al., 1989) whereas use rates are at least equivalent for both age groups in the studies employing a whole-list format (with the exception of Schneider & Uhl, 1990, and when performance goals were set in West & Thorn, 2001). Because parts of the lists can be revisited, the whole-list format eliminates the need to maintain previously presented words for clustering, making it less demanding. General cognitive resources should therefore be less important for successful clustering in the whole-list format as opposed to the individual-words format. In particular, WMC, which involves an individual's ability to maintain information while simultaneously conducting processing, should be a more necessary resource for clustering in an individual-words than whole-list format (cf., Wegesin et al., 2000). Given the strong age-related declines in WMC (Bopp & Verhaeghen, 2005), semantic clustering should be very taxing on older adults in an individual-words format, resulting in reduced clustering production, particularly under dual-task demands. However, older adults' reduced WMC should not notably limit their clustering use in the whole-list format. Hence, list presentation format may moderate the occurrence of age-related production deficiencies for semantic clustering by affecting the demands of clustering on general cognitive resources, particularly WMC.

Little is known about influences on older adults' successful utilization of semantic clustering. Taconnat et al. (2009) report a lack of clustering benefits in older adults in an individual-words format. Most other studies employing this format do not report the clustering – recall correlations but Witte et al. (1993) report a strong positive correlation in older adults and Amrhein et al. (1999) report a positive correlation for their whole sample. So far, research has not examined the effects of semantic clustering instructions without pairing them with physical sorting of words on index cards (Basden, Basden, & Bartlett, 1993; Guttentag, 1988; Kliegel, Altgassen, Martin, & Kruse, 2003; Worden & Meggison, 1984). That is, it is unclear if older adults could generally benefit from semantic clustering instructions without further support and if such benefits would be influenced by presentation format. Like a production deficiency, a utilization deficiency should be more likely in the more resource-demanding individual-words format than in the whole-list format.

Metacognitive Aspects

The fact that older adults produce encoding strategies under certain conditions but not others suggests no general deficit in older adult's strategy knowledge (cf., Bottiroli et al., 2010). However, age-related differences in the metacognitive appreciation of encoding strategy costs and benefits may still contribute to age-related differences in encoding strategy choices. Generally, metacognitive monitoring appears to be spared from negative age-related effects and both younger and older adults demonstrate similarly good metacognitive awareness of differential strategy effectiveness (Hertzog & Hultsch, 2000). However, some age-related differences have been reported for older adults' monitoring of performance aspects of strategy use (e.g., of response time benefits; Hertzog, Touron, & Hines, 2007). Paired with higher cognitive costs of strategy utilization, impaired monitoring of strategy benefits would render older adults even less likely to use an encoding strategy. Importantly, age-related differences have been reported for memory control beliefs with older adults being less likely than younger adults to believe that they can control their own memory performance (Lachman, 2006). Some studies found that such differences in control beliefs partially contribute to age-related production deficiencies for encoding strategies (Hertzog, McGuire, & Lineweaver, 1998; Lachman & Andreoletti, 2006) whereas others did not find such a relation (Blatt-Eisengart & Lachman, 2004; Hertzog, McGuire, Horhota, & Jopp, 2010). One methodological problem might be that memory beliefs were primarily assessed in a general manner rather than specific to the particular memory task. Given that age-related production deficiencies only occur in some task settings, probing such beliefs in more specific task contexts may be important, especially given the evidence that older adults' strategy use appears sensitive to task factors. For example, older adults' use of a memory-based strategy in skill acquisition tasks is more closely related to memory confidence for that specific task than general memory confidence (Touron & Hertzog, 2004a). Thereby, older adults' awareness of differential demands of strategy use in different tasks may play a key role in effects of task factors on their strategy production. Given that people's metacognitive judgments are often initially insensitive to external, not item-inherent, factors until more experience is gained (Koriat, 1997; Touron, Hertzog, & Speagle, 2010), older adults might sometimes fail to benefit from a task's particular affordability of an encoding strategy due to lacking metacognitive awareness. Therefore, it is important to consider metacognitive beliefs about strategy benefits and costs alongside their objective assessments.

Goals and Hypotheses of the Present Study

The goal of the present study was to examine the influence of presentation format (individual words vs. whole list) on older adults' production and utilization of semantic clustering. For the first time, presentation format was manipulated (between participants) within an experiment, keeping other task factors (e.g., word material and study duration) constant. Of particular interest was whether list presentation format moderates the occurrence of an age-related production deficiency for semantic clustering, explaining inconsistent prior research findings. In particular, older adults were expected to spontaneously cluster less than younger adults in the individual-words format but to a similar extent in the whole-list format. The whole-list format was predicted to afford semantic clustering use by making it less demanding on general cognitive resources, particularly WMC, than in the individual-words format. Further, cognitive costs of semantic clustering, assessed more directly via secondary task interference (cf., Naveh-Benjamin et al., 2005), were expected to depend on general cognitive resources and thus to be higher in the individual-words format as well as in older adults. Cognitive costs were expected to be positively correlated with spontaneous clustering use, such that individuals experiencing higher cognitive costs would be less likely to spontaneously use clustering. Thereby, age-related differences in semantic clustering use in the individual-words format were expected to be mediated by the age-related increase in cognitive costs of clustering use due to older adults' reduced general cognitive resources. Additionally, semantic clustering and recall performance after instructions to use clustering were assessed, to determine whether the presentation-format differences on clustering resource demands would also translate into more successful clustering utilization in the whole-list format, especially for older adults. Participants' metacognitive beliefs about clustering difficulty and efficacy were also probed, to provide a subjective assessment of cognitive costs and benefits of semantic clustering.

CHAPTER II

METHOD

Participants

Younger adult participants were 132 undergraduates (age range 18 – 26 years) from the University of North Carolina at Greensboro who received course credit. Nine additional younger adults were tested but replaced due to having previously participated in a pilot study with the same material (two), difficulty understanding English (one), being older than 30 (two), and experimenter error resulting in repeated presentation of a word list (two). Older adults were recruited from the community via newspaper ads and talks at senior centers and received a $25 gift card to a local department store for completing the 2.5 hour testing session. Several older participants had previously participated in studies in our laboratory but none of these involved semantic clustering or working memory tasks. Older adults with neurological disorders or taking medications that affect memory were excluded from participating (based on self-reports during scheduling). A total of 120 older adults were included in the analysis (age range 60 – 81 years). Two additional older adults were tested but replaced because complete tone task data were missing due to an experimenter error. One additional older adult who produced almost exclusively intrusions during recall was replaced. All participants had at least 20/50 corrected near-visual acuity. Sample characteristics will be presented in the Results section.

Design

Throughout the session each participant studied three categorizable lists with the first one assessing spontaneous semantic clustering use (no strategy instructions) while for the other two lists use of semantic clustering was instructed at study, first under full attention (FA) and then under dual task while simultaneously classifying tones (DT). Presentation format of the lists was manipulated between participants within each age group such that for half of the participants all lists were presented in an individual-words format with one word appearing at a time in the center of the screen. For the other half, a whole-list format was used such that all study list words were presented simultaneously on the screen in two columns of 10 words each. Younger and older adults were randomly assigned to presentation-format conditions with an equal number of younger (66; 15 males) and older (60; 15 males) adults per condition. Thus, the design was a 3 (List: spontaneous vs. FA clustering instructed vs. DT clustering instructed) x 2 (Age Group: younger vs. older adults) x 2 (Presentation Format: individual words vs. whole list) mixed factorial.

Material and Procedure

All computerized tasks were run in E-Prime 2.0 Professional (Psychology Software Tools Inc., Pittsburgh, PA), recording response times in milliseconds.

Word list study and recall. Three categorizable word lists containing five words each from four distinct semantic categories (i.e., total of 20 words per list) were constructed based on recent category norms (Van Overschelde, Rawson, & Dunlosky, 2004). The lists are in Appendix A. For each category, exemplars named by at least 20%

but no more than 80% of the norming sample in response to the category label were selected with no differences in mean naming percentage between lists (List A: $M = 0.49$, $SE = .04$; List B: $M = 0.48$, $SE = .04$; List C: $M = 0.49$, $SE = .03$), all ts<1, nor in the mean number of syllables (maximum of three syllables; List A: $M = 1.65$, $SE = .15$; List B: $M = 1.65$, $SE = .17$; List C: $M = 1.70$, $SE = .16$), all ts<1. Each participant studied each of the three lists. For the first list, participants were simply instructed to study the words for later recall with no mention of encoding strategies or the words being categorizable, allowing assessment of spontaneous strategy use (i.e., spontaneous list). For the second and third list, instructions mentioned benefits of semantic grouping and participants were asked to group the words by category during study. For the second list, participants had full attention during encoding (i.e., FA clustering-instructed list) whereas the third list was encoded under dual-task demands with simultaneous completion of a tone-discrimination task described below (i.e., DT clustering-instructed list). Participants never completed a secondary task during recall. Which list served in which position was completely counterbalanced in each condition.

List presentation during study varied by presentation format condition with all three lists being presented in the same format to a given participant. In the individual-words conditions, each word was presented centered on the computer screen for 3s with a 500ms fixation cross between words. For the whole-list conditions, all 20 list words were presented at once on the screen - in two columns of 10 words each - for 69.5 s. Thereby, the total study time was held constant across presentation-format conditions.[1] In both formats, the order of words was randomized for each participant with the restriction that at least two words intervened between words (in the same column for the whole-list condition) from the same semantic category. After studying a list, participants were asked to continuously deduct three from a provided three-digit number for 30 s, entering each result. Then participants were instructed to type any words they remembered from the study list. Importantly, test instructions did not mention anything about the order in which the words should be typed, even when semantic clustering had been instructed. Participants saw all words they had already typed on the screen. Three minutes were given for recall, timed by the computer with no option to end early.

After recall of each list, participants answered several questions related to strategy use. For the first spontaneous list, participants indicated any strategies used on a check list immediately after recall. For all lists, participants indicated their use of semantic clustering during study and test on a percentage scale (0-100% in 10% increments), the difficulty of semantic clustering use during study and test (on a 5-point scale from ‘not at all difficult’ to ‘very difficult’), and their beliefs about the effectiveness of semantic clustering during study and test for their own memory and memory of other people their age (on a 5-point scale from ‘not at all’ to ‘very much’). See Appendix B for the exact wording of these questions. Then participants were asked to freely name any categories the study words could be grouped into and were then probed with the actual list

[1] Study time was equated including fixation-cross presentations in between words in the individual-words format because participants likely use this extra time for further processing of the just presented word as well as rehearsal.

categories and asked to type any study words they remembered from a given category (category order was randomized for each participant). There was no time limit for this category-cued recall test.

Tone-discrimination task. A high-pitched (2,000 Hz) and a low-pitched (200 Hz) tone of 150 ms duration were recorded using the NCH Tone Generator software (http://www.nch.com.au/tonegen/index.html). Participants used headphones with the volume adjusted to their own preference. On the keyboard the up-arrow key was labeled as 'H' for 'high' and the down-arrow key as 'L' for 'low.' Participants were instructed to classify the tones using these keys as quickly and accurately as possible. Tones were presented in blocks of 24 with 12 high- and 12 low-pitched tones with six of each in the first and six each in the second half in a random order. The time between tones was randomly sampled (without replacement) from a fixed list of 24 durations varying between 2,000 and 4,000 ms. By using a list of fixed durations but randomizing the order for each block the total block time was held constant at 69.5 s (i.e., the study time for the word lists) but participants could not predict when the next tone would occur. Participants first completed practice trials with speed and accuracy feedback. After the first practice block of 24 tones, participants completed additional practice blocks until 91.7% accuracy (no more than two errors) was reached on a block (nonresponses counted as inaccurate). The majority of participants reached this criterion in the second practice block. Two older adults (one in each presentation format condition) needed one additional practice block and one older adult in the individual-words format needed three additional practice blocks.

Next, participants completed one more block of 24 tones without feedback to assess baseline speed for making tone discriminations alone. If no response was made to three tones a reminder was displayed on the bottom of the computer screen. If a participant failed to respond to more than six tone presentations, the block was restarted with a reminder to respond to all tones. Then participants were instructed that they would be asked to study and group a list of words while doing tone discriminations and in order to prepare for this challenge they would first practice reading information on the screen while classifying tones. In this block of 24 tone presentations a list of 20 pronounceable non-words was presented on the screen in the individual-words or whole-list format depending on the condition. Participants were instructed to (silently) read these non-words while responding to the tones as quickly and as accurately as possible with emphasis on the equal importance of both tasks. No feedback was provided but the reminder was displayed and the list was reset with a reminder if participants failed to respond to more than six tones. Finally, the tone discrimination task was added to the study phase of the third word list (see above). Instructions emphasized the importance of doing both, responding to all tone presentations as quickly and as accurately as possible and to study and group the words from the list. Instructions explicitly stated that some slowing down on the tone discriminations due to the encoding demands was normal and to be expected. During the study phase, reminders were displayed on the bottom of the screen if there was no response to three tones but the study list was not reset in the case of multiple nonresponses.

Vocabulary knowledge. Vocabulary knowledge was assessed with a computer-adapted version of the *ETS Advanced Vocabulary Test I* (Version 4; Ekstrom, French, & Harman, 1979). Participants selected the synonym for 36 difficult target words from five options per target. For each set of 18 synonym problems participants had 4 minutes with an option to end early. Participants could move back and forth between items as wanted, make changes to previous answers within a set, and skip items. Guessing was discouraged. The final score was the total number of correctly selected synonyms (maximum 36).

Processing speed tasks. Processing speed was assessed with two timed paper-pen tasks. For each, the experimenter timed with a stopwatch, giving simultaneous ‘start’ and ‘stop’ signals to all participants in a session. In the *Digit-Symbol Substitution Task* (Wechsler, 1981), participants were presented with nine digit-symbol pairings and copied as many symbols as possible in 90s for a random digit sequence. The final score was the total number of correctly copied symbols (maximum 100). In the *Pattern Comparison Task* (Salthouse, 1996), participants were presented with two line patterns and had to decide if the patterns were the same or different by writing ‘S’ or ‘D’ next to them. The task contained two letter-sized sheets each containing 30 patterns in two columns. Participants had 30s per sheet. The final score was the total number of correct comparisons across both pages (maximum 60).

WMC tasks. Two computerized automated complex span tasks (Kane et al., 2004; Unsworth, Heitz, Schrock, & Engle, 2005) were used to assess WMC. In the verbal *Reading Span Task*, participants judged whether a sentence was plausible or not and after each sentence a letter was presented for 1 s. After two to six sentence-letter pairs, participants selected the letters in the order they had been presented from a list of letters with the option to leave blanks for forgotten letters. Participants first practiced each part (sentence judgments, letter recall) separately and combined before completing three sets of each size (2 - 6) in a randomly mixed order for a total of 48 trials. Importantly, each participant's average judgment latency during the 15 sentence-judgment practice trials was recorded and if sentence judgments during the real trials exceeded this average latency by more than 2.5 standard deviations the judgment was timed out (counted as an error) and the next letter was shown. Participants were informed about this time constraint and instructed to always make their sentence judgment as quickly as possible while maintaining accuracy. The purpose of this time constraint is to ensure that participants stay on task and do not devote time to rehearsal or strategic remembering (see Unsworth et al. for more details). In addition, participants received feedback about the number of letters correctly recalled as well as their sentence judgment accuracy after each trial and were asked to maintain at least 85% accuracy on the sentence judgments. The final reading span score was the total number of letters recalled in the correct position (maximum 60). In the spatial *Symmetry Span Task*, the processing component was judging whether patterns in an 8x8 square matrix are symmetric along the vertical axis. For the memory component, participants then saw a 4x4 square matrix with one square filled in red for 650 ms. After sets of two to five symmetry judgments and square presentations, participants had to select the previously presented red squares in the 4x4 matrix in the order they had been presented in. Again, each task component was practiced and mean latency on the 15 practice symmetry judgments plus 2.5 standard deviations was used as the time limit for symmetry judgments during the real trials. Feedback about performance was given after each trial and participants were asked to maintain 85% accuracy on the symmetry judgments. The real trials comprised three sets of each size (2 - 5) in a randomly mixed order for a total of 42 trials. The final symmetry span score was the total number of squares recalled in the correct position (maximum 42).

Metacognitive beliefs. In addition to self-reports of strategy use, clustering difficulty, and task-specific clustering effectiveness beliefs on each study list (see earlier), participants completed three computerized scales from the *Metacognition in Adulthood (MIA)* questionnaire (Dixon & Hultsch, 1983). On the strategy subscale, participants indicated how frequently (5-point scale from 'never' to 'always') they use internal (e.g., imagery) as well as external (e.g., calendar) memory strategies in their daily lives (18 items). On the locus subscale, participants indicated via their level of agreement (5-point scale from 'strongly disagree' – 'strongly agree') with nine statements to what extent they believe memory performance can be internally controlled. On the achievement subscale, participants indicated how important doing well on memory tasks is to them via their agreement with 16 statements (same 5-point scale as locus). Items from the three subscales were randomly intermixed. For each scale, each participant's average rating across items was calculated.

General procedure and task order. Participants were tested in age-homogenous groups of up to six in a room with individually separated computer stations with an experimenter present throughout the entire session. Participants first signed a consent form and took a near (corrected) visual-acuity test. Tasks were then completed in the same following order with breaks offered between tasks as necessary: (1) Digit-symbol substitution, (2) pattern comparison, (3) first spontaneous word list, (4) vocabulary test, (5) reading span task, (6) FA clustering-instructed list, (7) tone-discrimination practice as well as performance of tone-discriminations alone and while reading non-words, (8) DT clustering-instructed list, (9) MIA, and (10) symmetry span task. Upon completion of all tasks, participants completed a computerized demographic questionnaire and were then debriefed and dismissed. Sessions lasted up to 2 hours for younger adult participants and up to 2.5 hours for older adult participants.

CHAPTER III

RESULTS

Sample Demographics and General Cognitive Abilities

Table 1 displays demographic information about the sample as well as performance on the vocabulary, processing speed and WMC tasks, and responses on metacognitive scales. Each measure was analyzed with a 2 (Age Group: younger vs. older adults) x 2 (Presentation Format: individual words vs. whole list). The statistics for all age group comparisons are given in Table 1 (as *t*-tests). Older adults had completed more years of formal education and took more medications than the younger adults but the age groups did not differ on self-rated health. For demographic variables, condition effects emerged only for formal years of education with a main effect of presentation format, $F(1, 248) = 6.71$, $MSE = 4.40$, $\eta_p^2 = .03$, $p = .010$, that was qualified by a significant interaction with age group, $F(1, 248) = 7.31$, $MSE = 4.40$, $\eta_p^2 = .03$, $p = .007$. The two younger adult conditions did not differ in average years of formal education, $t<1$, but the older adults in the individual-words format condition had on average completed one more year of formal schooling than their peers in the whole-list condition, a significant difference, $t(118) = 2.86$, $d = 0.36$, $p = .005$. Importantly, the range of education years[2] was equivalent for the two conditions (12 – 20 years) with all older

[2] One older adult in the whole-list condition indicated only having five years of formal schooling but having completed a High School Diploma so the completed years were adjusted to 12.

adults having at least completed a High School Diploma and there were no condition differences in vocabulary performance, *Fs* $\leq$ 1.93, $p \geq .166$, indicating equivalent verbal knowledge of the older adult conditions despite the education difference. Expectedly, older adults outperformed younger adults on vocabulary knowledge.

Younger adults outperformed older adults on the processing speed tasks, with no main effect of presentation format or interaction of age group and presentation format, all *F*s<1. For WMC tasks, at least 70% average accuracy was required on the processing component (i.e., sentence or symmetry judgments) to avoid trade-offs between the processing and storage component of the tasks, resulting in reduced sample sizes as indicated in Table 1.[3] As expected, younger adults outperformed older adults on both WMC tasks. Further, there were significant differences between the presentation format conditions on both span tasks, $F(1, 246) = 10.14$, $MSE = 88.9$, $\eta_p^2 = .04$, $p = .002$, for reading span, and $F(1, 223) = 6.04$, $MSE = 50.5$, $\eta_p^2 = .03$, $p = .015$, that did not interact with age group, both F<1. In both age groups, the individual-words condition scored higher than the whole-list condition on the WMC tasks. This difference was unexpected given the random assignment to presentation-format conditions and is particularly

[3] Typically, a processing accuracy criterion of 85% is used (e.g., Unsworth et al., 2005). In the current study, 16.7% of the younger and 30.5% of the older adults performed below 85% accuracy on the symmetry judgments. This may be due to fatigue because symmetry span was the last task tested. Importantly, there was no evidence for trade-off (i.e., negative correlation between processing accuracy and span scores). Rather, processing accuracy and span scores were positively correlated in the younger adults, $r(132) = .360$, $p < .011$, for reading span, and $r(132) = .375$, $p < .001$ for symmetry span, as well as for symmetry span in the older adults, $r(120) = .282$, $p = .002$, with no correlation for reading span in the older adults, $r(120) = .126$, $p = .169$. Therefore, a more lenient processing criterion of 70% accuracy on the processing task was used (chance performance would be 50%).

surprising given no condition differences in processing speed, a related indicator of general cognitive ability. Notably, processing speed tasks were administered first in the testing session whereas the WMC tasks were administered later after other tasks. Thus, presentation format may have influenced WMC task performance. After studying in a whole-list format, participants may have had difficulty adapting to the more fast-paced presentation format of the WMC tasks that was more akin to the individual-words format. Because task order was constant for all participants any such influences should be similar across individuals within a presentation-format condition; hence within each condition the WMC tasks should still tap into meaningful individual differences even though the across-condition comparisons of WMC performance may not be meaningful. For both processing speed and WMC composite scores were derived by *z*-transforming task scores using the whole sample's (i.e., aggregating across age groups and presentation format conditions) mean and standard deviation and averaging *z*-scores for the two processing speed tasks and for the two WMC tasks, respectively.[4] Expectedly, processing speed and WMC composite scores were positively correlated, $r(131) = .210$, $p = .016$, for younger adults and $r(120) = .479$, $p < .001$, for older adults.

[4] If a participant only met the processing criterion for one WMC task, that tasks' *z*-score was used for the composite to retain a larger number of participants for analysis. Performance on the two span tasks was significantly correlated, $r(131) = .389$, $p < .001$, for younger adults and $r(120) = .454$, $p < .001$, for older adults. Only the reading span task *z*-score was used as the WMC composite score for five younger adults in the whole-list format, ten older adults in the individual-words format and nine older adults in the whole-list format. Further, only the symmetry span task *z*-score was used for one older adult in the individual-words format. One younger adult in the individual-words format did not meet the processing criterion on either span task. Any reported correlations with WMC were similar when using a stricter composite score requiring scores from both WMC tasks as well as when not applying a processing criterion.

Table 1

Sample Demographics and Performance on Vocabulary, Processing Speed, and Working Memory Span Tasks

Measure	Younger Adults: Individual Words	Younger Adults: Whole List	Older Adults: Individual Words	Older Adults: Whole List	Age-Related Difference
			Demographic Information		
Age (years)	19.61 (0.21)	19.71 (0.25)	68.72 (0.57)	70.28 (0.62)	
Formal Education (years)	13.48 (0.17)	13.52 (0.19)	16.33 (0.29)	15.05 (0.38)	$t(250) = 8.49$, $d = 1.07$, $p < .001$
Subjective Health Rating	0.59 (0.08)	0.70 (0.08)	0.57 (0.08)	0.75 (0.10)	$t(250) = 0.17$, $d = 0.02$, $p = .867$
Number Medications	0.67 (0.13)	0.50 (0.11)	2.50 (0.24)	2.87 (0.24)	$t(250) = 11.19$, $d = 1.41$, $p < .001$
			General Cognitive Abilities		
Vocabulary	14.80 (0.42)	13.70 (0.45)	22.72 (0.91)	21.87 (0.95)	$t(250) = 11.42$, $d = 1.44$, $p < .001$
Pattern Comparison	39.09 (1.02)	40.13 (0.84)	28.81 (0.69)	28.77 (0.65)	$t(250) = 13.24$, $d = 1.67$, $p < .001$
Digit-Symbol Substitution	66.17 (1.36)	67.22 (1.26)	52.37 (1.22)	52.42 (1.56)	$t(250) = 10.77$, $d = 1.36$, $p < .001$
Reading Span	49.25 (0.91) n = 65	45.74 (1.16) n = 66	45.02 (1.20) n = 59	40.92 (1.48) n = 60	$t(248) = 3.74$, $d = 0.47$, $p < .001$
Symmetry Span	28.49 (0.84) n = 65	26.39 (0.99) n = 61	17.02 (0.94) n = 50	14.45 (1.00) n = 51	$t(225) = 12.30$, $d = 1.64$, $p < .001$
			Metamemory in Adulthood Questionnaire Scales		
Achievement	3.93 (0.05)	3.85 (0.05)	4.03 (0.05)	4.00 (0.04)	$t(249) = 2.79$, $d = 0.35$, $p = .006$
Locus	3.63 (0.06)	3.55 (0.05)	3.60 (0.08)	3.65 (0.06)	$t(249) = 0.60$, $d = 0.08$, $p = .552$
Strategy – internal	3.83 (0.06)	3.80 (0.05)	3.57 (0.07)	3.51 (0.07)	$t(249) = 4.52$, $d = 0.57$, $p < .001$
Strategy – external	3.60 (0.07)	3.67 (0.07)	4.03 (0.08)	3.78 (0.08)	$t(249) = 3.66$, $d = 0.46$, $p < .001$

Note. Standard errors in parentheses. Subjective health rating is on 5-point Likert scale comparing to a perfect status of health (0 = 'very good'; 4 = 'very poor'). Maximum scores are 36 for vocabulary (Ekstrom et al., 1979), 60 for pattern comparison (Salthouse, 1996), 100 for digit-symbol substitution (Wechsler, 1981), 60 for reading and 42 for symmetry span (Unsworth et al., 2005; 70% accuracy required on processing component hence the reduced number of scores). Metamemory in Adulthood (MIA; Dixon & Hultsch, 1983) scores are average responses on 5-point Likert scale (1 = 'strongly disagree'; 5 = 'strongly agree'). One older adult in the whole list condition did not complete the symmetry span task and the MIA questionnaire due to running out of time.

Regarding the MIA scales (data missing for one older adult in the whole-list condition), older adults scored higher on memory achievement than younger adults but the age groups did not differ on locus beliefs. For all conditions, mean achievement and locus scores were above the scale midpoint (2.5) indicating that participants generally were motivated to perform well on memory tasks and believed that memory could be internally controlled. Typical age-related patterns occurred with regards to frequency of memory strategy use in daily life with older adults reporting less frequent use of internal (encoding) strategies but increased use of external strategies (e.g., Bouazzaoui et al., 2010; Loewen, Shaw, & Craik, 1990). For achievement, locus, and internal memory strategies there were no effects involving presentation format, all $F \leq 1.47$. For external strategy use, there was no main effect of presentation format, $F=1.51$, but an interaction of age group and presentation format, $F(1, 247) = 4.55$, $MSE = .35$, $\eta_p^2 = .02$, $p = .034$. The younger adult groups did not differ, $t<1$, but older adults in the individual-words condition reported using external memory strategies more frequently than their peers in the whole-list condition, $t(117) = 2.32$, $d = 0.43$, $p = .022$.

Main Analysis Overview

The present study's design was a 3 x 2 x 2 mixed factorial with list (spontaneous vs. FA clustering instructed vs. DT clustering instructed) as within-subjects factor and age group (younger vs. older adults) and presentation format (individual words vs. whole list) as between-subjects factors. Therefore, unless otherwise noted for a particular measure, dependent variables were analyzed with a 3 x 2 x 2 mixed analysis of variance (ANOVA). The sphericity assumption across the different lists was tested using

Mauchley's *W* statistic for all dependent measures and, if violated, Greenhouse-Geisser corrected *df*s were used. Alpha was set to .05. For *F*-tests, effect sizes are reported as partial eta squared (η_p^2) with .01 conventionally indicating a small effect, .06 a medium effect, and .14 a large effect (Cohen, 1988). For follow-up pairwise comparisons, Cohen's *d* (Cohen, 1988) was computed with the web-based calculator at http://www.cognitiveflexibility.org/effectsize/, using the pooled standard deviation for between-subjects comparisons and Morris and DeShon's (2002) equation 8 for within-subjects comparisons. Cohen's suggested conventions for *d* are 0.2 for a small effect, 0.5 for a medium effect, and 0.8 for a large effect. Pearson *r* was used for correlations involving continuous measures and Spearman *rho* for correlations with self-reports given on ordinal Likert scales. Given the multitude of variables assessed, only key correlations of interest will be presented here but the full correlation matrices for all conditions are presented in Appendix C.

Recall Performance

Typed words were matched to study-list words by the first four letters, which was the smallest set of letters that uniquely identified each study word. To further minimize the influence of spelling errors, words not matching with a study word on the first four letters were manually checked. Both the singular and the plural of a study word were accepted (e.g., 'sock' or 'socks'). Sometimes participants repeated words in the recall output, often to correct a prior misspelling. The first occurrence of a word in the recall output was kept and any later repetitions were deleted to ensure that recall output order represented the order in which words first came to mind. The following analyses will focus on the proportion of words freely recalled but more detailed analyses of semantic aspects of the recall output (e.g., number of categories recalled and words recalled per category) and the additional category-cued recall are presented in Appendix E.

For each participant, the proportion of words recalled from each of the three 20-word lists was computed. Mean performance is displayed in Figure 1. Recall performance greatly differed across lists, $F(1.88, 464.71) = 143.22$, $MSE = .02$, $\eta_p^2 = .37$, $p < .001$. Compared to the spontaneous list, recall was significantly better on the FA clustering-instructed list, $t(251) = 7.18$, $d = 0.45$, $p < .001$, but significantly worse on the DT clustering-instructed list, $t(251) = 8.25$, $d = 0.52$, $p < .001$. Regarding age-related differences there expectedly was a large main effect of age group, $F(1, 248) = 88.99$, $MSE = .05$, $\eta_p^2 = .26$, $p < .001$, with older adults recalling fewer words than younger adults. This effect of age group interacted with list, $F(1.88, 466.73) = 10.15$, $MSE = .02$, $\eta_p^2 = .04$, $p < .001$. The age-related decline in recall performance was evident on all three lists but varied in magnitude with a smaller age-related difference on the spontaneous list, $t(250) = 3.95$, $d = 0.50$, $p < .001$, compared to when clustering use was instructed, FA $t(250) = 8.46$, $d = 1.07$, $p < .001$, and DT $t(250) = 8.50$, $d = 1.07$, $p < .001$. That is, age-related recall differences magnified after instructions to use semantic clustering and this magnification persisted under divided attention.

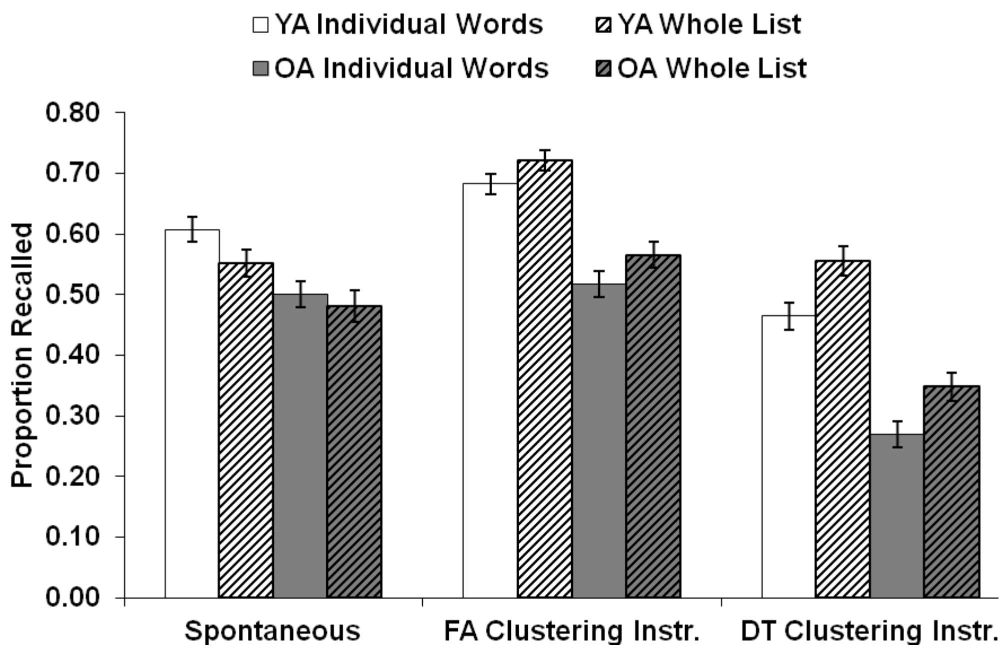

Figure 1. Mean proportion recalled by age group, presentation format, and list. Error bars represent standard errors. YA = younger adults; OA = older adults; FA= full attention; DT = dual task; Instr. = instructed.

Regarding effects of presentation format, the main effect was marginally significant, $F(1, 248) = 3.67$, $MSE = .05$, $\eta_p^2 = .02$, $p = .056$, and significantly interacted with list, $F(1.88, 464.71) = 12.11$, $MSE = .02$, $\eta_p^2 = .05$, $p < .001$. Whereas there was no recall difference between the two presentation formats on the first spontaneous list, $t(250) = 1.67$, $d = 0.21$, $p = .097$, the whole-list conditions recalled significantly more words once clustering use was instructed, both under FA, $t(250) = 2.06$, $d = 0.26$, $p = .040$, and, especially, under DT $t(250) = 3.25$, $d = 0.41$, $p < .001$. This is in contrast to the marginal trend on the spontaneous list where the whole-list conditions tentatively recalled fewer words than the individual-words conditions. Thus, this presentation-format effect cannot be attributed to generally better memory of participants in the whole-list conditions but rather the whole-list format appears to specifically benefit recall when semantic clustering is used as the primary encoding strategy (as instructed). There were no interactions involving age group and presentation format, both $F < 1$, suggesting that these beneficial presentation-format effects were similar across age groups. It is noteworthy that for older adults in the individual-words condition the clustering instructions did not lead to a significant increase in average recall from the spontaneous to the FA clustering-instructed list, $t < 1$, reflecting the combination of both the reduced clustering-instruction benefit in older adults and the reduced clustering-instruction benefit in the individual-words format. The increase was significant in the older adults whole-list condition and both younger adult conditions, all $p < .05$.

Semantic Clustering

Relative output clustering. Several indices have been proposed to quantify the amount of semantic clustering in a participant's recall output (e.g., Bousfield, 1953; Frender & Doubilet, 1974; Roenker, Thompson, & Brown, 1971). These indices are based on the number of category repetitions (i.e., adjacent recall of two words from the same semantic category) observed in the recall output, adjusting in different ways for the total number of words recalled. It has been argued that the adjusted ratio of clustering (ARC) score varies the least with other aspects of recall (Roenker et al., 1971) but in the present data, ARC score distributions strongly deviated from normality (cf., Taconnat et al., 2009) and were incomputable for many recall outputs, especially on the DT list. Instead, the *ratio-of-repetition* (RR) score (Bousfield, 1953; Frender & Doubilet, 1974) was used. RR is computed as

$$RR = \frac{R}{n-1} \qquad (1)$$

where R is the total number of category repetitions observed in the recall output and n is the total number of items recalled. RR scores vary between 0 and 1 with 0 indicating the absence of semantic clustering and 1 indicating perfect clustering within an output (i.e., category repetitions only). Importantly, RR scores are independent of the total number of items recalled (Murphy, 1979) and can be computed as long as at least two words were recalled. RR is often used in developmental work including cognitive aging (e.g., Amrhein et al., 1999; West et al., 2009; West & Thorn, 2001; Witte et al., 1993). RR score distributions sufficiently approximated normality, allowing for powerful parametric analysis. As detailed in Appendix D, RR and ARC scores were highly correlated and showed similar mean patterns. Hence, the choice of RR does not influence the conclusions drawn.

Due to recall of only one or no word, RR scores were incomputable for seven older adults in the individual-words format condition (all for DT list) and five older adults in the whole-list condition (one for FA list, four for DT list), who were excluded from analysis.[5] Mean RR scores are presented in Figure 2A.

[5] Since high clustering should result in better recall, incomputable clustering scores due to low recall are more likely to be low scores and are hence not missing at random, violating assumptions for a mixed-model analysis that would allow considering all available data. Importantly, a mixed-model analysis run for explorative purposes on RR as well as LBC yielded the same pattern of results. That is, the reported results were not biased by exclusion of participants with missing scores. Another approach considered was to set missing RR and LBC scores to 0 which also let to the same conclusions but it is questionable if low or zero recall necessarily implies the absence of any clustering.

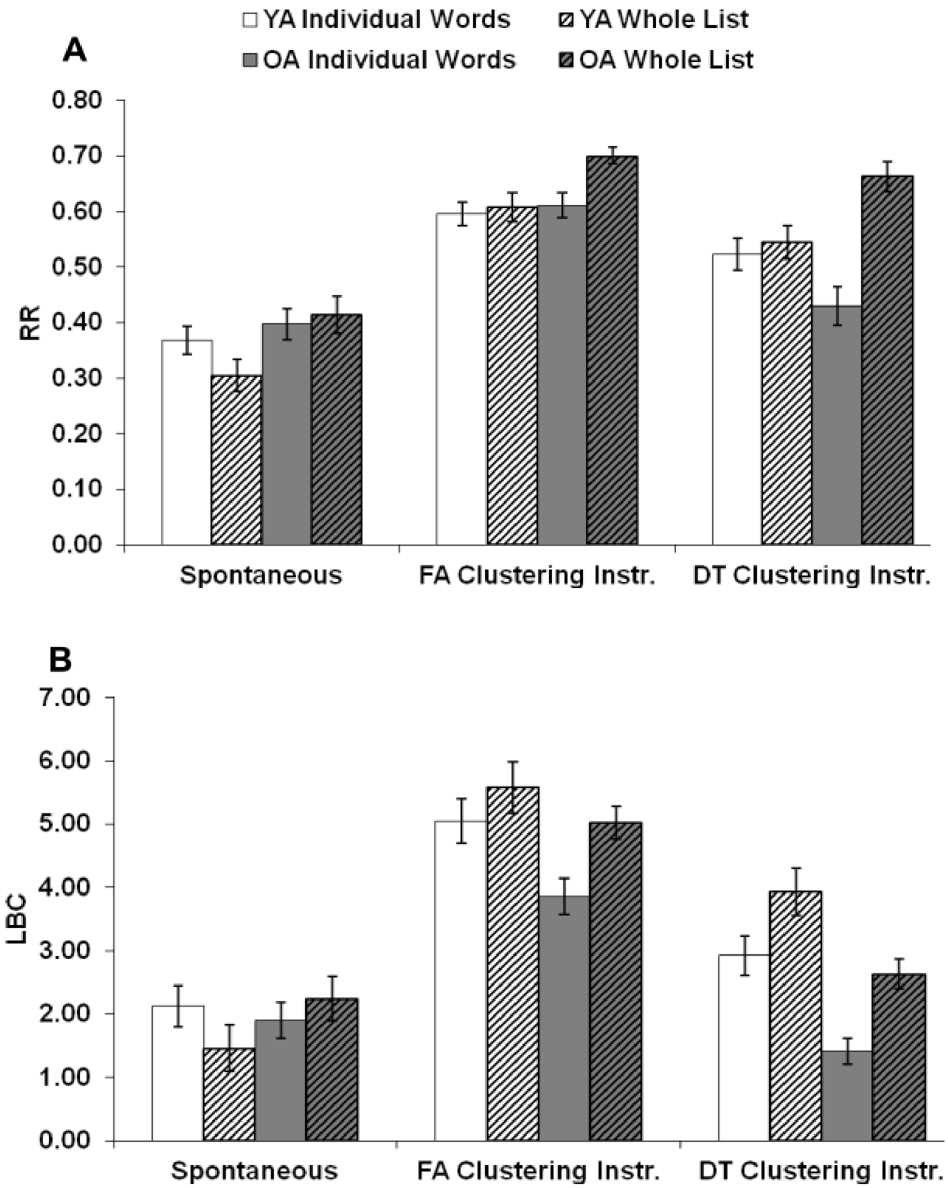

Figure 2. Mean clustering indices by age group, presentation format, and list. Part A shows the ratio of repetition (RR; Bousfield, 1953) reflecting relative output clustering and part B the list-based clustering index (LBC; Stricker et al., 2002) reflecting absolute clustering success. Error bars represent standard errors. FA = full attention; DT = dual task; Instr. = instructed.

Figure 2 displays data only from participants included in the analysis (i.e., with computable scores on all three lists).

There was a large main effect of list, $F(1.84, 433.00) = 112.97$, $MSE = .04$, $\eta_p^2 = .32$, $p < .001$. Compared to the spontaneous list, output clustering was higher when the use of semantic clustering was instructed at study, especially under full attention, $t(239) = 15.81$, $d = 1.03$, p < .001, but also under divided attention, $t(239) = 8.48$, $d = 0.55$, $p < .001$. Output clustering on the FA clustering-instructed list was higher than on the DT clustering-instructed list, $t(239) = 5.39$, $d = 0.31$, $p < .001$. That is, the instructions to group words by category during study successfully increased semantic clustering in the recall outputs, even under divided attention. Counter to predictions, older adults overall clustered slightly more within their recall outputs, $F(1, 236) = 5.67$, $MSE = .06$, $\eta_p^2 = .02$, $p = .018$. Age group did not interact with list, $F(1.84, 433.00) = 1.44$, $p = .237$. There was further a main effect of presentation format, $F(1, 236) = 7.38$, $MSE = .06$, $\eta_p^2 = .03$, $p = .007$, that was qualified by an interaction with list, $F(1.84, 433.00) = 9.49$, $\eta_p^2 = .04$, $p < .001$. RR scores did not differ between the two presentation-format conditions on the spontaneous list, $t<1$, but once clustering was instructed within-output clustering was greater in the whole-list format both under FA, $t(238) = 2.07$, $d = 0.27$, $p = .039$, and, especially, under DT, t(238) = 3.81, $d = 0.49$, $p < .001$. Importantly, the interaction of age group and presentation format was significant, $F(1, 236) = 10.60$, $MSE = .06$, $\eta_p^2 = .04$, $p = .001$. Separate ANOVAs by age group revealed an overall effect of presentation format in the older adults, $F(1, 106) = 21.35$, $MSE = .05$, $\eta_p^2 = .17$, $p < .001$, but not in the younger adults, $F<1$. The three-way interaction was only marginally significant, $F(1.84, 433.00) = 2.46$, $MSE = .04$, $\eta_p^2 = .01$, $p = .091$, but it is evident from the significant List x Presentation Format interaction that the format differences in RR in older adults were not present on the first spontaneous list, t<1, but rather when clustering was instructed, FA $t(106) = 3.39$, $d = 0.65$, $p = .001$, and DT $t(106) = 5.34$, $d = 1.03$, $p < .001$ (all $t \leq 1.66$, $p \geq .099$ for younger adults). Importantly, given these selective effects of presentation format in the older adults, age-related differences in RR varied across presentation format. There was no age-related difference in the individual-words format, F<1, whereas older adults showed significantly greater output clustering than younger adults in the whole-list format, $F(1, 119) = 14.94$, $MSE = .07$, $\eta_p^2 = .11$, $p < .001$. Interestingly, in the individual-words format age group and list interacted, $F(1.90, 222.17) = 3.77$, $MSE = .04$, $\eta_p^2 = .03$, $p = .027$, with no age-related differences on the spontaneous and FA clustering-instructed list, both t<1, but older adults having *lower* output clustering on the DT clustering-instructed list, $t(117) = 2.10$, $d = 0.39$, $p = .038$. No such interaction occurred in the whole-list format, F<1, indicating that older adults benefitted greatly from the whole-list format in terms of being able to produce highly clustered recall outputs.

Absolute clustering success. RR is a recall-based clustering index which describes how much the recall output was clustered. Thereby, it is relative to how many words were recalled. In other words, high clustering of just a few recalled words results in a similarly high score as high clustering of many recalled words. Alternatively, a *list-based* clustering (LBC) index (Stricker, Brown, Wixted, Baldo, & Delis, 2002) can be computed as

$$LBC = R - \frac{[(n-1)(m-1)]}{N_L - 1} \qquad (2)$$

where R is again the number of category repetitions observed in the recall output, n the total number of words recalled, m the number of members per category on the list (i.e., 5 in the present study) and N_L the total number of items on the study list (i.e., 20). The fraction of the equation yields the number of category repetitions that would be expected based on random sampling of the number of recalled items from the study list at recall (i.e., chance-expectation for clustering). Importantly, unlike recall-based indices, which only consider what was recalled (i.e., the output), the LBC index considers properties of the original study list, reflecting the assumption that semantic clustering occurs during encoding of the study list. Thereby, the LBC index rewards greater numbers of beyond-chance semantic clusters and longer clusters of same-category words (see Stricker et al. for exemplary output protocols). Consequently, LBC is not fully independent of recall with higher scores for greater recall if recall is clustered. This index has been used in cognitive aging research (Fenandes & Grady, 2008; Gross & Rebok, 2011; Jacobs et al., 2001; Wegesin et al., 2000), including research on dementia (Delis et al., 2010). Complementary to the RR analysis, which described relative output clustering, the LBC index provides valuable information about absolute clustering success. The two indices were highly correlated, $.429 \leq r \geq .969$, all $p < .05$ (see Appendix C).

LBC is computable as long as at least one word is recalled, resulting in missing scores and exclusion from analyses for only two older adults in the individual-words condition and one older adult in the whole-list condition who recalled nothing on the DT list (see Footnote 5). See Figure 2B for means. There was a large main effect of list, $F(1.92, 471.39) = 134.78$, $MSE = 4.44$, $\eta_p^2 = .36$, $p < .001$. Compared to the spontaneous list, LBC scores significantly increased with clustering instructions, especially under FA, $t(248) = 15.09$, $d = 0.96$, $p < .001$, but still under DT, $t(248) = 4.10$, $d = 0.27$, $p < .001$. Expectedly, LBC was lower on the DT clustering-instructed list compared to the FA clustering-instructed list, $t(248) = 12.85$, $d = 0.82$, $p < .001$.

There was a significant main effect of age group, $F(1, 245) = 7.51$, $MSE = 10.93$, $\eta_p^2 = .03$, $p = .007$, that further interacted with list, $F(1.92, 471.39) = 10.75$, $MSE = 4.44$, $\eta_p^2 = .04$, $p < .001$. There was no age-related difference on the first spontaneous list, $t< 1$, but younger adults had higher LBC scores when clustering was instructed, FA $t(247) = 2.56$, $d = 0.32$, $p = .011$, and DT $t(247) = 4.58$, $d = 0.58$, $p < .001$. Thus, after instructions to semantically cluster, older adults clustered less successfully compared to younger adults, especially under DT, paralleling the findings of larger age-related recall differences on the clustering-instructed lists.

Regarding effects of presentation format, the main effect was significant, $F(1, 245) = 6.12$, $MSE = 10.93$, $\eta_p^2 = .03$, $p = .014$, and interacted with list, $F(1.92, 471.39) = 6.63$, $MSE = 4.44$, $\eta_p^2 = .03$, $p = .002$. The presentation format conditions did not differ on the first spontaneous list, $t<1$, but once clustering was instructed the whole-list conditions clustered more successfully, FA $t(247) = 2.44$, $d = 0.31$, $p = .015$, and DT $t(247) = 3.57$, $d = 0.45$, $p < .001$. Descriptively, the presentation-format difference in LBC appears to be somewhat more strongly pronounced in the older adults, paralleling the strong presentation-format effect on older adults' RR scores, but given the greater variability of the LBC measure this interaction was not significant, $F(1, 245) = 1.60$, $p = .208$, nor was the three-way interaction, $F<1$. Importantly, the LBC measure reveals a benefit of the whole-list format in the younger adults when clustering was instructed, which was not evident in the RR scores but is in line with the greater overall recall in the whole-list format found in both age groups on the FA and DT lists.

Self-reported use of semantic clustering. After studying and recalling each list, participants indicated which percentage of the study list words they had semantically clustered during study and test (from 0% - 100% in increments of 10%; see Appendix B). Individual estimates for study and test clustering use were very similar, $.852 \leq r(252) \leq .916$, all $p < .001$, and were averaged for analysis; separate analyses yielded similar patterns. Descriptive statistics are in Table 2. There was a main effect of list, $F(1.58, 392.23) = 134.78$, $MSE = 759.32$, $\eta_p^2 = .30$, $p < .001$. Compared to the spontaneous list, participants reported clustering on a significantly greater proportion of the study list after clustering instructions both under FA, $t(251) = 14.64$, $d = 0.94$, $p < .001$, and under DT, $t(251) = 7.72$, $d = 0.49$, $p < .001$. Clustering self-reports were lower under DT than FA, $t(251) = 6.43$, $d = 0.42$, $p < .001$.

There was a main effect of age group, $F(1, 248) = 5.57$, $MSE = 1862.36$, $\eta_p^2 = .02$, $p = .019$, that interacted with list, $F(1.58, 392.23) = 4.93$, $MSE = 795.32$, $\eta_p^2 = .02$, $p = .013$. There were no age-related differences in reported clustering on the spontaneous list, t<1, but once clustering was instructed younger adults reported clustering a greater proportion of the list words than older adults, FA $t(250) = 3.38$, $d = 0.60$, $p = .001$, and DT $t(250) = 2.64$, $d = 0.47$, $p = .009$. This pattern parallels the age-related differences in LBC scores on the FA and DT list but not the RR scores which were higher for older adults in the whole-list condition on these lists. Indeed, participants were asked to indicate what percentage of the study-list words (not just of the words recalled) had been clustered; hence it should be more akin to LBC than RR. Thus, older adults were sensitive to their reduced clustering success; even though older adults in the whole-list condition produced highly clustered recall outputs they were aware that they had clustered a relatively small portion of the study list. Compared to the free-recall proportion, however, both younger and older adults' estimates were quite overconfident.

Table 2

Means and Standard Errors for Self-Reports About Use, Difficulty, and Efficacy of Semantic Clustering

	Younger Adults		Older Adults	
List	Individual Words	Whole List	Individual Words	Whole List
	Self-Reported Semantic Clustering Use (%)			
Spontaneous	52.80 (3.98)	38.56 (4.44)	46.58 (4.24)	46.17 (4.81)
FA Clustering Instructed	85.00 (2.49)	82.42 (3.17)	70.50 (4.14)	73.92 (3.80)
DT Clustering Instructed	71.59 (3.76)	73.79 (3.93)	58.67 (5.12)	63.83 (4.58)
	Self-Reported Beliefs about Semantic Clustering Efficacy			
Spontaneous	3.81 (0.11)	3.71 (0.10)	3.87 (0.11)	3.72 (0.10)
FA Clustering Instructed	4.06 (0.11)	4.22 (0.11)	3.93 (0.11)	3.91 (0.11)
DT Clustering Instructed	3.78 (0.12)	3.77 (0.12)	3.73 (0.13)	3.75 (0.11)
	Self-Reported Semantic Clustering Difficulty			
Spontaneous	2.12 (0.10)	2.16 (0.11)	2.39 (0.10)	2.33 (0.13)
FA Clustering Instructed	1.90 (0.11)	1.90 (0.11)	2.59 (0.15)	2.21 (0.11)
DT Clustering Instructed	3.07 (0.16)	2.81 (0.15)	3.72 (0.16)	3.08 (0.16)

Note. Standard errors in parentheses. Clustering use was reported on a 11-point percentage scale (0-100% in 10% increments). Difficulty and efficacy were reported on 5-point Likert scales (1 = 'not at all difficult'/'not at all effective' to 5 = 'very difficult'/'very effective'). For each measure, reports for study and test were averaged. FA = full attention; DT = dual task.

Regarding presentation-format effects, there was no main effect or interaction with age group, both *F*s ≤ 1.46. The interaction of presentation format with list was marginally significant, $F(1.58, 392.23) = 3.20$, $MSE = 759.32$, $\eta_p^2 = .01$, $p = .054$, but pairwise comparisons were not significant for any of the lists, all $t(251) \leq 1.75$, $p \geq .081$, with the largest numerical difference occurring on the spontaneous list due to the fairly low clustering reports of younger adults in the whole-list format. The three-way interaction was not significant, $F<1$. That is, neither younger nor older adults were sensitive to the presentation-format effects with higher clustering success in the whole-list conditions. Within conditions, however, self-reports were sensitive to the achieved clustering as evidenced by significant correlations with both RR (not significant for older adults whole-list condition for FA and DT lists; all other $.268 \leq r \leq .724$, $p < .05$) and LBC ($.273 \leq r \leq .714$, all $p < .05$; see Appendix C).

Relation between Semantic Clustering and Recall Performance

A beneficial effect of clustering-instructions on recall is evident from the observed increase in proportion recalled between the spontaneous and the FA clustering-instructed list reported earlier. Additionally, correlational analyses reveal if individual differences in recall are accompanied by individual differences in the level of semantic organization. Table 3 presents the correlations between all clustering measures and recall, which were generally positive and moderate to strong, indicating that increased organization of the study words by semantic category was related to recall of more study words. With the exception of the FA and DT clustering-instructed list in the whole-list conditions, correlations were never significantly smaller in older compared to younger adults, all $p \geq .347$, and even trended to be higher in some cases ($p = .093$ for the age-group comparison in the individual-words conditions on the FA list). That is, even though the clustering-instruction benefit was reduced in the older adults, the clustering-recall relationship was similar across age groups on an individual level. This was even true for the older adult individual-words condition where recall did not significantly improve with the clustering instructions. It appears that some older adults in this condition were able to cluster quite successfully despite the lack of benefit on the group-level. Notable exceptions are the nonsignificant correlations between RR and recall on the clustering-instructed lists in the older adult whole-list condition. This condition achieved the highest mean RR scores on the FA and DT list and low RR scores occurred rarely; that is, individual differences in clustering were less pronounced. Further, this finding nicely demonstrates that the RR score is independent of the level of recall. For example, 12 older adults in this condition (20%) recalled only words from one category on the DT list, resulting in the maximum RR score (1.0) but a low proportion recalled. Generally, however, participants in this condition clustered quite successfully, as evident in their increased LBC scores which were strongly correlated with recall.

Table 3

Correlations between Clustering Indices and Proportion Recalled

	Younger Adults		Older Adults	
Clustering Index	Individual Words	Whole List	Individual Words	Whole List
		Spontaneous List		
RR	.543*	.607*	.491*	.484*
LBC	.625*	.610*	.682*	.713*
Self-Report	.546*	.598*	.626*	.674*
		FA Clustering-Instructed List		
RR	.373*	.426*	.419*	-.043
LBC	.662*	.649*	.851*	.868*
Self-Report	.241+	.619*	.503*	.509*
		DT Clustering-Instructed List		
RR	.383*	.592*	.612*	-.107
LBC	.792*	.741*	.829*	.806*
Self-Report	.508*	.593*	.590*	.470*

Note. Peason *r* is reported for RR and LBC but Spearman *rho* for the self-reports that were given on a 10%-increments percentage scale (averaged for study and test). RR = ratio-of-repetition score (Bousfield, 1953). LBC = list-based semantic clustering index (Stricker et al., 2002).
* $p < .05$. + $.05 > p < .10$.

Mediation of age-group differences in recall through semantic clustering. Mediation analyses were conducted to examine if age-related differences in recall were mediated through semantic clustering following Baron and Kenny (1986). Since there were no age-related differences in either clustering index on the spontaneous list, age-related differences in spontaneous recall cannot be mediated through semantic clustering. Further, given the lack of age-related reductions in relative output clustering (RR), only clustering success (LBC) is a potential mediator of the magnified age-related differences in recall on the clustering-instructed lists.[6] Path coefficients were estimated using SPSS AMOS and are displayed in Figure 3.

[6] An exception was the DT clustering-instructed list where RR was significantly lower in older compared to younger adults in the individual-words format. Analyses revealed

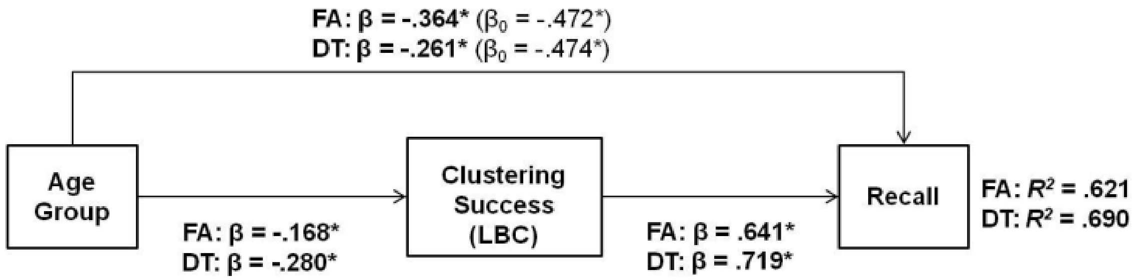

Figure 3. Mediation of age-group differences in recall through semantic clustering success. β_0 indicates the zero-order correlation. Age group is coded as 0 = younger and 1 = older adults. FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; LBC = list-based clustering index (Stricker et al., 2002).
* $p < .05$.

On both the FA and DT list the negative age-group effect on recall was reduced by additional inclusion of LBC as a predictor. Sobel's *z* test indicated that there was significant mediation of the age-group effect through clustering success, $z = 2.67$, $p = .008$, for the FA list, and $z = 4.46$, $p < .001$ (also confirmed in confidence intervals based on 1000 bootstrap samples with the bias-corrected percentile method; cf. Preacher & Hayes, 2008). Importantly, on both lists age group predicted recall beyond LBC, indicating only partial mediation. That is, age-related recall differences after clustering instructions were in part due to older adults' lower clustering success.

Mediation of presentation format differences in recall through semantic clustering. Similarly, it was examined whether the differences between the presentation-format conditions (i.e., greater recall in the whole-list conditions) on the clustering-instructed lists were mediated by semantic clustering. The RR measure differed between presentation format conditions only in older but not in younger adults, whereas there

significant mediation in this case but because both clustering indices were highly correlated (compare Appendix C) and because age-related differences were more consistent in the LBC measure, LBC was chosen to be presented as a mediator.

were consistent presentation-format differences in LBC across all participants. Therefore, LBC was examined as a potential mediator, with the relevant regression coefficients displayed in Figure 4. The presentation format effects became non-significant when LBC was included as a predictor and Sobel's *z* test indicated significant mediation on both lists, $z = 2.50$, $p = .013$, for the FA list and $z = 3.52$, $p < .001$, for the DT list. That is, presentation-format differences in recall on the clustering-instructed lists were fully mediated by the format differences in clustering performance. In other words, the whole-list format conditions achieved higher recall on the clustering-instructed lists because participants were able to cluster more successfully on these lists.

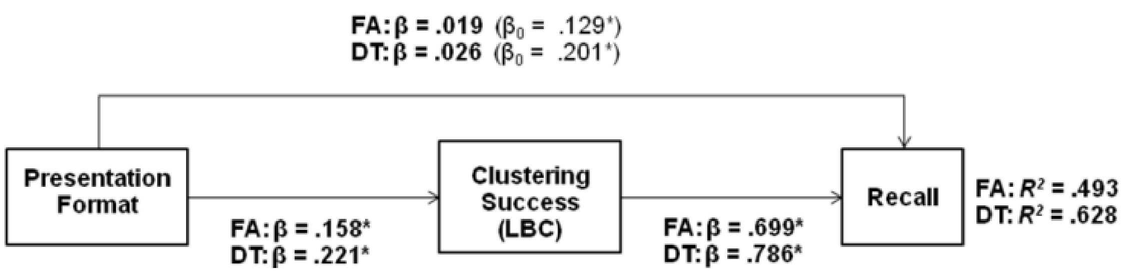

Figure 4. Mediation of presentation-format differences in recall through semantic clustering success. β_0 indicates the zero-order correlation. Age group is coded as 0 = younger and 1 = older adults. FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; LBC = list-based clustering index (Stricker et al., 2002).
* $p < .05$.

Beliefs about clustering efficacy. Participants also reported their beliefs about the efficacy of semantic clustering during study and test for improving one's own memory and memory of same-aged peers for a list like the one just studied. Separate analyses of these reports yielded similar patterns; for simplicity, reports were averaged for each participant to represent general clustering efficacy beliefs. Descriptive statistics are presented in Table 2. There was a main effect of list, $F(1.92, 475.32) = 16.30$, $MSE = .36$, $\eta_p^2 = .06$, $p < .001$, that interacted with age group, $F(1.92, 475.32) = 3.11$, $MSE = .36$, $\eta_p^2 = .01$, $p < .047$. For younger adults, efficacy beliefs were highest after the FA clustering-instructed list, $t(131) = 4.93$, $d = 0.43$, $p < .001$, compared to the spontaneous list, and $t(131) = 5.97$, $d = 0.44$, $p < .001$, compared to the DT clustering-instructed list. Efficacy estimates for the spontaneous and DT list did not differ, $t<1$. For older adults, the same general pattern emerged but the increase in efficacy beliefs after the FA clustering-instructed list was less pronounced. Thereby, efficacy estimates for the FA clustering-instructed list were only significantly different from estimates for the DT clustering-instructed list, $t(119) = 2.70$, $d = 0.25$, $p = .008$, but only marginally compared to the spontaneous list, $t(119) = 1.84$, $d = 0.17$, $p = .069$. Again, estimates for the spontaneous and DT list did not differ, $t<1$. Notably, mean efficacy estimates on the spontaneous list were above the mid-point of the scale (2.5) indicating that participants believed semantic clustering to be a beneficial strategy but after experiencing clustering benefits on the FA clustering-instructed list these estimates rose even higher. The reduced increase in older adults' efficacy beliefs parallels the reduced increase in recall after clustering instructions. As a consequence, older adults' efficacy estimates were marginally lower compared to the younger adults on the FA clustering-instructed list, $t(250) = 1.94$, $d = 0.25$, $p = .053$, with no such tendencies for an age-related difference on the other two lists, both $t<1$ and $F<1$ for the main effect of age group. No other effects were significant, all $F \leq 1.84$. That is, there were no effects of presentation format even though presentation format had affected recall and clustering success, paralleling participants' insensitivity to presentation format in the self-reports of clustering use.

Cognitive Costs or Difficulty of Semantic Clustering

Tone-discrimination data from one older adult in the individual-words condition was excluded due to very low tone-discrimination accuracy; it might be that this participant switched keys once passing the practice phase accuracy criterion because accuracy was below chance. As evident in Table 4, all other participants responded very accurately to nearly all tone presentations, even when simultaneously clustering study words.

Tone-discrimination RTs. Response times (RTs) from inaccurate responses or faster than 200 ms (0.4% of all accurate RTs) were excluded from analysis.[7] For each participant and each processing phase (alone vs. non-words vs. DT clustering) median RTs were computed to reduce the influence of individual outlying RTs. Means of these median RTs are presented in Table 4 and were analyzed with a 3 (Processing: alone vs. non-words vs. clustering) x 2 (Age Group: younger vs. older) x 2 (Presentation Format: individual-words vs. whole-list) mixed ANOVA, using Greenhouse- Geisser corrected *df*.

[7] Tone-discrimination response proportions and accuracy were near ceiling, leaving a sufficient number of accurate tone responses for RT observation for all participants. Formal analyses revealed that response proportions decreased with additional processing $F(1.70, 420.31) = 43.81$, $MSE = .001$, $\eta_p^2 = .15$, $p < .001$ (clustering < non-words < alone, all $p < .05$, all $d \leq 0.55$). Age group interacted with processing, $F(1.70, 420.31) = 17.44$, $MSE = .001$, $\eta_p^2 = .07$, $p < .001$, with no age-related differences in the alone phase, $t(249) = 1.48$, $p = .142$, but older adults making slightly fewer responses during non-words, $t(249) = 3.28$, $d = 0.41$, $p = .001$, and, especially, during clustering, $t(249) = 5.94$, $d = 0.75$, $p < .001$. For accuracy, there was only a significant processing effect, $F(1.93, 476.04) = 11.60$, $MSE = .001$, $\eta_p^2 = .05$, $p < .001$ (clustering < non-words < alone, all $p < .05$, all $d \leq 0.29$).

The main effect of processing was significant, $F(1.69, 417.14) = 673.68$, $MSE = 17606.26$, $\eta_p^2 = .73$, $p < .001$. Reading non-words slowed down tone RTs compared to performing the tone discriminations alone, $t(250) = 25.97$, $d = 1.96$, $p < .001$. Clustering slowed RTs beyond RTs observed during non-word reading, $t(250) = 14.30$, $d = 0.96$, $p < .001$. That is, as expected, semantic clustering carried cognitive costs.

Table 4

Tone Discrimination Performance

Processing	Younger Adults: Individual Words	Younger Adults: Whole List	Older Adults: Individual Words	Older Adults: Whole List
	% Tone Responses			
Alone	99.9 (.06)	100 (.00)	99.9 (.10)	99.9 (.10)
Non-words	98.9 (.35)	99.4 (.29)	97.5 (.60)	97.8 (.56)
DT Clustering	99.0 (.31)	99.0 (.43)	96.4 (.73)	94.4 (.82)
	% Accurate Responses			
Alone	98.8 (.25)	98.5 (.34)	98.5 (.33)	97.9 (.45)
Non-words	97.6 (.48)	97.6 (.40)	98.2 (.45)	97.6 (.52)
DT Clustering	97.9 (.43)	97.2 (.48)	96.5 (.56)	96.4 (.69)
	Mean Response Times (ms)			
Alone	410 (8.54)	402 (7.39)	532 (13.05)	512 (13.40)
Non-words	577 (14.43)	610 (17.12)	803 (24.96)	792 (22.21)
DT Clustering	691 (17.57)	778 (25.13)	962 (33.11)	1019 (32.57)

Note. Standard errors are in parentheses. Mean response times are averages of individual median response times for accurate tone discriminations that were at least 200 ms.

Expectedly, older adults were slower than younger adults, $F(1, 247) = 142.27$, $MSE = 48539.58$, $\eta_p^2 = .37$, $p < .001$, and this age group effect further interacted with processing, $F(1.69, 417.14) = 21.20$, $MSE = 17606.26$, $\eta_p^2 = .08$, $p < .001$. To follow up on this interaction, absolute cost measures were computed as RT difference scores. The first difference scores concerns the slowing of RTs by adding non-word reading to the tone task. Older adults were more slowed ($M = 275$, $SE = 13.17$) by this additional processing than younger adults (M = 187, SE = 9.92), $t(249) = 5.23$, d = 0.66, $p < .001$. Of most interest to the present study is the difference in RTs between the non-word reading and the DT clustering, which reflects costs of semantic clustering use and is plotted in Figure 5. Older adults had greater clustering costs, that is older adults were slowed more by the added clustering task compared to the non-word reading than younger adults, $t(249) = 2.30$, $d = 0.29$, $p = .022$.

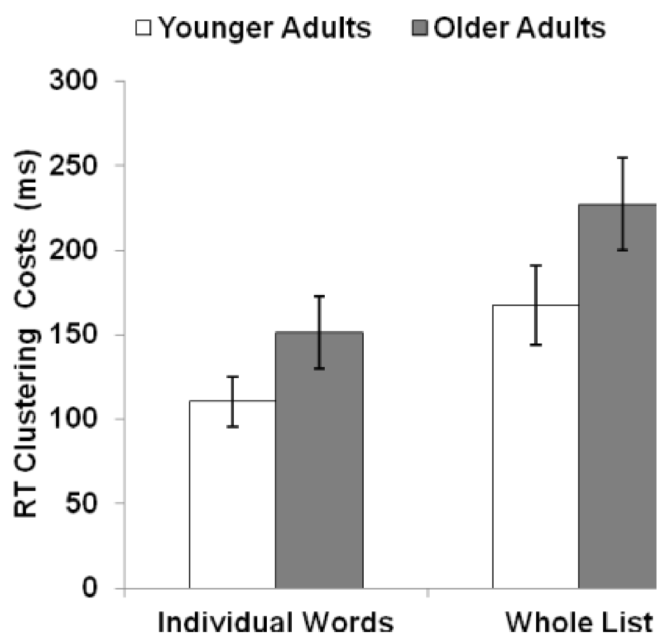

Figure 5. Response time (RT) clustering costs by age group and presentation format. Clustering costs were computed as the differences between tone-discrimination RTs while clustering and RTs while reading non-words. Error bars represent standard errors.

Regarding presentation format effects, there was no main effect, $F = 1.96$, $p = .163$, but an interaction with processing, $F(1.69, 417.14) = 8.24$, $MSE = 17606.26$, $\eta_p^2 = .03$, $p = .001$. There were no RT differences between the presentation format conditions when the tone discrimination task was performed alone, $t(249) = 1.07$, $p = .284$, or while reading non-words, $t<1$, but the whole-list conditions responded to the tones significantly slower while clustering, $t(249) = -2.32$, $d = 0.29$, $p = .021$. Hence, counter to expectations, the whole-list conditions had a higher cost of semantic clustering use (see also Figure 5). No other effects were significant; that is, the presentation format effect did not interact with age group, both $F<1$. That is, older adults had generally higher absolute costs of semantic clustering use, independent of the presentation format.

It has been suggested that costs should be considered relative to baseline performance to account for age-related differences in baseline performance (cf., Somberg & Salthouse, 1982). Thus, the absolute clustering costs were divided by non-word RTs and these relative clustering costs were analyzed with a 2 (Age Group: younger vs. older adults) x 2 (Presentation Format: individual words vs. whole list) univariate ANOVA. There were no age-related differences in this relative clustering cost measure, $F<1$, but the presentation-format effect was still present, $F(1, 247) = 6.08$, $MSE = .07$, $\eta_p^2 = .02$, $p = .014$, with no interaction, $F<1$. That is, younger and older adults did not differ in relative clustering costs but relative clustering costs were higher for the whole-list format conditions (younger: $M = .29$, $SE = .04$; older: $M = .31$, $SE = .04$) than for the individual-words conditions (younger: $M = .22$, $SE = .03$; older: $M = .22$, $SE = .03$).

In summary, RTs were sensitive to processing added to the tone-discrimination task. In particular, there was a significant increase in tone-discrimination RTs when participants were asked to simultaneously cluster study words compared to when they were merely asked to read non-words appearing on the screen. The absolute semantic clustering cost was significantly larger in older than younger adults but this age-related difference was not evident when costs were considered relative to baseline (non-word) RTs. Further, the semantic clustering cost was higher in the whole-list conditions (both measured absolute and relative), counter to the hypothesis that this condition would facilitate clustering use.

Correlates of RT clustering costs. Table 5 presents correlations between RT clustering costs and recall as well as clustering performance on the DT clustering-instructed list. Notably, recall and clustering performance trended to be positively correlated with clustering costs but only significantly in the whole-list conditions. That is, participants with better memory and categorical organization showed greater interference with tone-discrimination RTs during encoding. This finding is important for interpretation of the above-presented clustering cost analyses. Given that the age groups and presentation format conditions achieved different levels of recall as well as clustering the clustering costs do not necessarily reflect difficulty of clustering so much as achieved processing during encoding. Given that the correlations were only significant in the whole-list format, conditions for an analysis of covariance (controlling for achieved processing) were not met. It thus unfortunately remains unclear how the clustering costs would compare for the two formats and age groups if a similar level of semantic processing of the study lists had been achieved.

Table 5

Correlates of Clustering Costs and Self-Reported Clustering Difficulty

Clustering Index	Younger Adults		Older Adults	
	Individual Words	Whole List	Individual Words	Whole List
	Correlation with RT Clustering Costs			
DT Proportion Recalled	.195	.417*	.189	.268*
DT RR	.144	.463*	.102	.171
DT LBC	.167	.414*	.086	.352*
Processing Speed	-.134	.081	.084	.139
WMC	.120	.135	.254+	.198
	Correlation with Self-Reported DT Clustering Difficulty			
DT Proportion Recalled	-.552*	-.192	-.442*	-.385*
DT RR	-.170	-.245*	-.189	.017
DT LBC	-.418*	-.202	-.286*	-.363*
Processing Speed	.060	-.142	-.012	.070
WMC	.124	.030	-.072	.091
RT Clustering Costs	-.026	.049	.060	-.128

Note. Peason r is reported for correlations with RT clustering costs but Spearman *rho* for the self-report of clustering difficulty (ordinal 5-point Likert scale). DT = dual-task list; RR = ratio-of-repetition score; LBC = list-based semantic clustering index; WMC = working-memory capacity.
$^{*}p \leq .05$. $^{+}.05 < p \leq .10$.

Table 5 further displays correlations of clustering costs with processing speed and WMC for which negative correlations had been predicted such that participants with reduced cognitive resources would have larger clustering costs. Counter to this prediction, none of the correlations reached statistical significance; for older adults in the individual-words format the correlation between clustering costs and WMC was marginally significant but in the positive direction. Consequently, conditions were not met for WMC to mediate age-related differences in absolute clustering costs.

Self-reported clustering difficulty. Clustering difficulty reports (1 = ‘not at all difficult’ to 5 = ‘very difficult’; see Appendix B) for study and test were very similar on each list, $.693 \leq r(252) \leq .810$, all $p < .001$, and were averaged for analysis. Descriptive statistics are in Table 2. There was a main effect of list, $F(1.87, 463.69) = 102.74$, $MSE = .83$, $\eta_p^2 = .29$, $p < .001$. Compared to the spontaneous list, participants reported clustering to be similarly difficult on the FA clustering-instructed list, $t(251) = 1.53$, $p = .126$, but to be significantly more difficult on the DT clustering-instructed list, $t(251) = 10.36$, $d = 0.67$, $p < .001$. Further, older adults generally reported clustering to be more difficult, $F(1, 248) = 17.65$, $MSE = 1.66$, $\eta_p^2 = .07$, $p < .001$, consistent with the larger absolute RT clustering costs. This age group effect did not change across lists, $F(1.87, 463.69) = 1.83$, $p = .165$.

The main effect of presentation format was significant, $F(1, 248) = 5.29$, $MSE = 1.66$, $\eta_p^2 = .02$, $p = .022$, with greater difficulty reported in the individual-words conditions. This format effect was qualified by a significant interaction of presentation format and list, $F(1.87, 463.69) = 3.89$, $MSE = .83$, $\eta_p^2 = .02$, $p = .024$. There was no effect of presentation format on the spontaneous list, $t<1$, and on the FA clustering-instructed list, $t(250) = 1.49$, $p = .139$, but on the DT clustering-instructed list difficulty estimates were significantly higher in the individual-words conditions, $t(250) = 2.75$, $d = 0.35$, $p = .006$. Descriptively, the presentation format effect was stronger (and tentatively already present on the FA clustering-instructed list) in the older adults but the interaction of age group and presentation format was not significant, $F(1, 248) = 2.31$, $p = .130$, nor was the three-way interaction, $F<1$. The lower difficulty estimate for the whole-list conditions on the DT clustering-instructed list counters the higher RT clustering costs found in these conditions. This outcome supports the earlier suggestion that the higher clustering costs in the whole list conditions likely arise from the greater processing achieved rather than arising from greater difficulty of clustering in this format.

Table 5 presents the correlations of self-reported clustering difficulty on the DT list with recall and clustering on the DT list. There were significant negative correlations with proportion recalled and LBC in all but the younger adult whole-list condition. That is, participants who had achieved little recall and clustering (in terms of success) rated clustering use under DT to be more difficult. Again, no significant correlations with general cognitive resources (processing speed and WMC) emerged, counter to predictions. Finally, difficulty estimates were not correlated to RT clustering costs. That is, participants' experienced clustering difficulty was not reflected in greater interference with tone-discriminations during the DT clustering phase.

Correlates of Semantic Clustering

One goal of the present study was to identify correlates of spontaneous clustering use. In particular, it was predicted that spontaneous clustering would be influenced by its cognitive costs with participants with higher cognitive resources, particularly WMC, and lower clustering-induced RT interference being more likely to spontaneously use clustering. Table 6 displays correlations between spontaneous clustering use (RR and LBC) and general cognitive abilities (i.e., processing speed and WMC), vocabulary, as well as cognitive cost measures (i.e., RT clustering costs and difficulty estimates) and beliefs about clustering efficacy (task-specific assessment for each list). For the ordinal Likert-scale self-report measures (i.e., difficulty and efficacy estimates) Spearman *rho* is given, for all other variables Pearson *r* was used.

Table 6

Correlations between General Cognitive Abilities, RT Clustering Costs, Metacognitive Beliefs, and Semantic Clustering Indices

	Younger Adults				Older Adults			
	Individual Words		Whole List		Individual Words		Whole List	
	RR	LBC	RR	LBC	RR	LBC	RR	LBC
	Spontaneous List							
Processing Speed	.061	.104	.151	.174	-.032	.054	-.095	.136
WMC	.151	.160	.045	.089	-.173	-.051	-.121	.000
Vocabulary	.062	.028	-.186	-.136	.245+	.248+	-.048	.046
RT Clustering Costs	-.109	-.126	.103	.128	.006	.151	.111	.166
Difficulty Judgment	-.236*	-.247+	-.373*	-.377*	-.199	-.292*	-.315*	-.451*
Efficacy Belief	.233+	.231+	.389*	.422*	.013	.125	.224+	.217+
	FA Clustering-Instructed List							
Processing Speed	.015	.079	.018	.042	.245+	.267*	.041	.293*
WMC	.212+	.211+	.165	.125	.375*	.414*	.071	.385*
Vocabulary	.044	.106	-.015	.019	.030	.238+	-.057	.264*
RT Clustering Costs	.264*	.234+	.323*	.323*	.303*	.192	.130	.192
Difficulty Judgment	-.025	-.049	-.128	-.200	-.211	-.404*	.047	-.252+
Efficacy Belief	.047	.070	.226+	.230+	.218+	.303*	.053	.421*
	DT Clustering-Instructed List							
Processing Speed	.293*	.070	.233+	.241+	.066	.171	-.263+	.139
WMC	.185	.228+	.079	.020	.201	.165	-.249+	.233+
Vocabulary	.011	.130	.000	.006	.142	.207	-.012	.129
RT Clustering Costs	.144	.167	.463*	.414*	.102	.086	.171	.352*
Difficulty Judgment	-.170	-.418*	-.245*	-.202	-.189	-.286*	.017	-.363*
Efficacy Belief	.278*	.189	.370*	.367*	.253+	.334*	.026	.116

Note. Spearman *rho* correlations were used for the difficulty judgment and efficacy belief (both on ordinal 5-point Likert scales). Pearson *r* was used for all other variables. FA = full attention; DT = dual task; WMC = working memory capacity.
* $p < .05$. $^{+}$ $.05 < p < .10$.

Spontaneous semantic clustering. Counter to predictions, spontaneous clustering use was not related to general cognitive resources or RT clustering costs. This lack of correlations is in line with the lack of age-related differences. Despite reduced cognitive resources and increased (absolute) clustering costs, older adults were not less likely to spontaneously use semantic clustering. Vocabulary performance was also not related to spontaneous clustering. Interestingly, self-reported clustering difficulty was negatively correlated with spontaneous clustering use –that is, participants who believed clustering on the spontaneous list to be more difficult were also less likely to have used it. Efficacy beliefs were tentatively positively related to spontaneous use but these correlations were only significant in the younger adult whole-list condition.

Semantic clustering after instructions. Additionally, correlations with semantic clustering on the clustering-instructed lists (FA and DT) were examined. Since participants were instructed to use clustering, these correlations are not informative about strategy choice but rather inform about influences on clustering ability. Correlations with difficulty beliefs mostly vanished on the FA list, indicating that participants complied with clustering instructions independent of their beliefs about clustering difficulty, but under DT difficulty beliefs negatively related to clustering use in all conditions. Some positive correlations between clustering efficacy beliefs and clustering indices persisted on the clustering-instructed lists. Further, some positive correlations emerged with RT clustering costs. For the DT list, these correlations were already presented, and suggest that participants clustering more successfully under DT showed greater costs. For the FA list, these correlations (primarily significant in younger adults) were somewhat surprising but suggest that participants who generally clustered more successfully after instructions tended to show greater clustering costs under DT.

Most interestingly, positive correlations of general cognitive resources with clustering indices emerged. In particular, processing speed and WMC were positively correlated with absolute clustering success (LBC) on the FA list for both older adult conditions and also with relative output clustering (RR) in the individual-words format.

For WMC, correlations with clustering indices approached .2 in the younger adults but were not significant (marginal in the individual-words format). Across younger and older adults, correlations of WMC and clustering success (LBC) were significant, $r(125) = .368$, $p < .001$, in the individual-words format (for relative output clustering [RR] $r(125) = .260$, $p = .003$; one younger adult missing due to no valid WMC score), and $r(126) = .237$, $p < .001$, in the whole-list format (for RR, $r(125) = -.002$, $p = .983$). Across age groups, the correlation of clustering success with processing speed was significant in the individual-words format, $r(126) = .262$, $p = .003$, but only marginal in the whole-list format, $r(126) = .165$, $p = .065$ (not significant for RR in either format). When examined as simultaneous predictors, only WMC uniquely predicted clustering success. Therefore, WMC was examined as a potential mediator of age-related differences in clustering success on the FA clustering-instructed list.

All possible paths (fully saturated model) were estimated with SPSS AMOS, as displayed in Figure 6. As before, significance of meditational (indirect) effects was assessed with Sobel's z test but also converged with bootstrapped confidence intervals (bias-corrected percentile method based on 1000 bootstrapped samples). In the individual-words condition, age-related declines in clustering success (LBC) were mediated by age-related declines in WMC, $z = 3.09$, $p = .002$, with no remaining direct effect of age group on clustering success. That is, older adults' less successful clustering (resulting in no recall improvement with clustering instructions in this format) was fully explained by their reduced WMC. WMC did not directly influence recall performance.

That is, WMC did not mediate age-related differences in recall beyond the mediation through clustering success.

Age group's total effect on FA recall was an average reduction by 16.3%, of which about 4.7% were indirect through WMC → LBC with a remaining direct effect of 11.6%. For the whole-list condition, WMC also fully mediated age-related declines in clustering success, $z = 2.23$, $p = .026$. Interestingly, WMC further directly influenced recall performance and thereby significantly mediated age-related declines in recall independent of clustering success in this format, $z = 2.38$, $p = .010$. Again, a direct negative effect of age group on FA recall remained. Of the average 15.6% reduction in FA recall, about 2.4% were mediated through WMC → LBC and 2.1% through WMC alone, with a remaining significant direct effect of 11.1%. That is, in both presentation formats, age-related reductions in FA clustering success were fully mediated by WMC but together WMC and LBC only mediated part of the age-related decrease in FA recall performance. The model fit well when all path coefficients were set equal across groups, $\chi^2(6) = 6.41$, $p = .379$, indicating that the two format conditions did not differ in terms of the relations among these variables (although the WMC → recall paths was only significant in the whole-list conditions).

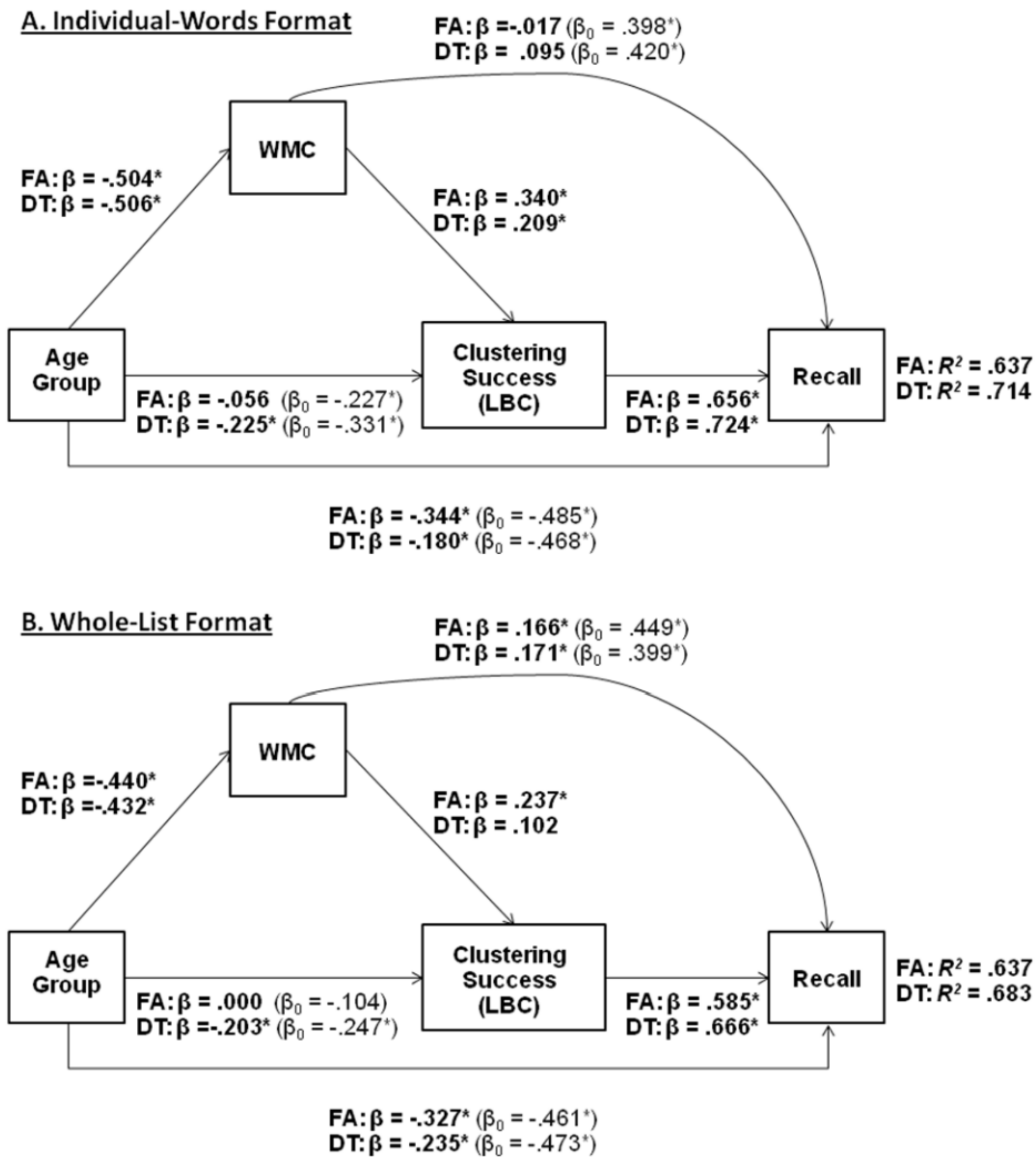

Figure 6. Path analysis with working memory capacity and clustering success as mediators of recall performance on the clustering-instructed lists. β_0 indicates zero-order correlations for mediated relationships. Age group is coded as 0 = younger adults and 1 = older adults. WMC = working memory capacity; LBC = list-based clustering index (Stricker et al., 2002); FA = full-attention clustering instructed list. DT = dual-task clustering-instructed list.
* $p < .05$.

Additionally, differences in intercepts in LBC and recall between format conditions were assessed. The LBC intercept was significantly higher in the whole-list conditions (5.44, *SE* = .33) than in the individual-words conditions (4.39, *SE* = .34), $\Delta\chi^2(1) = 4.87$, $p = .027$, confirming the prior finding of higher clustering success in the whole-list conditions but, importantly, when adjusting for WMC.[8] To further illustrate these differences, the WMC- LBC relationship for the two formats is contrasted in Figure 7. As is evident, those with lower WMC were able to cluster more successfully in the whole list compared to the individual-words format. In particular, older adults with low WMC had LBC scores close to chance (0) in the individual-words format but were still able to cluster above chance in the whole list format. Notably, the recall intercepts did not differ between the formats when controlling for WMC and clustering success (individual words: .47, *SE* = .02; whole list: .51, *SE* = .02), $\Delta\chi^2(1) = 2.03$, $p = .154$ (compare also Figure 4).

On the DT clustering-instructed list, correlations between WMC and clustering success (LBC) were significant across age groups, $r(123) = .323$, $p < .001$, in the individual-words conditions and $r(125) = .189$, $p = .034$, in the whole-list conditions (two older adults in the individual-words format and one in the whole-list format excluded due to incomputable LBC score), although none of the correlations within each cell reached significance. Given the significant overall correlation, it was also examined if WMC

[8] The intercepts for WMC differed significantly given the higher average WMC performance in the individual-words conditions. Because this WMC differences might not indicate actual ability differences but rather influences of the manipulation on the WMC task performance, intercept comparison analyses for both the FA and DT list were rerun to confirm that the differences in adjusted LBC means were still present even when WMC *z* scores were computed within each presentation format condition only.

mediates age-related differences in clustering success and recall on this list, see Figure 6 for DT path coefficients. In the individual-words format, WMC significantly mediated the age-related reduction in clustering success, $z = 2.03$, $p = .042$, but a significant direct effect of age group on clustering success remained indicating only partial mediation. WMC again did not directly influence DT recall performance. Of the 18.2% reduction in DT recall in older adults, 11.2% were indirect through WMC $\rightarrow$ LBC and the remaining direct effect of 7.0% was significant. In the whole-list format, there was no significant mediation of age-related reductions in DT clustering success through WMC, $z = 1.04$, $p = .289$, but WMC directly influenced recall performance and thereby mediated the age-related reduction in DT recall, $z = 2.61$, $p = .009$. Independently, clustering success (LBC) mediated the age-related reduction in DT recall, $z = 2.08$, $p = .037$. Of the 20.1% reduction in DT recall in older adults, 3.2% were mediated through WMC and 5.8% through LBC, leaving a significant direct effect of 10.0%). Again, path coefficients could be set equal across the format conditions with good model fit, $\chi^2(6) = 4.99$, $p = .545$, despite the format differences regarding the significance of meditational effects. The LBC intercept was significantly higher in the whole-list than in the individual-words format even when adjusting for WMC (Figure 7; see also Footnote 8), $\Delta\chi^2(1) = 9.30$, $p = .002$. Finally, the whole-list conditions achieved greater DT recall when adjusting for both WMC and clustering success, $\Delta\chi^2(1) = 6.11$, $p = .013$.

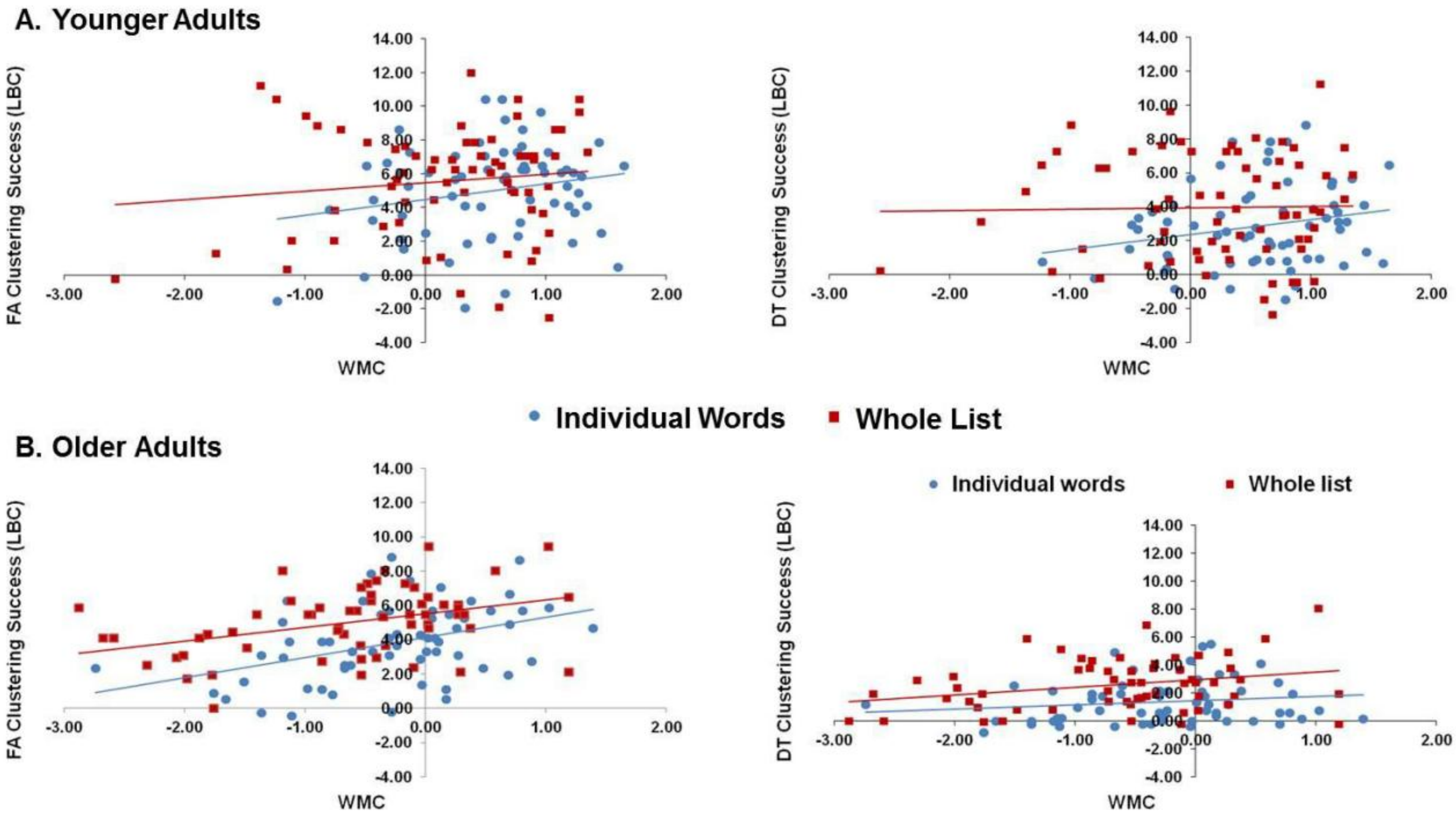

Figure 7. Correlation between working memory capacity and clustering success on the clustering-instructed lists. The upper two graphs show correlations for younger adults, the lower two for older adults. Lines represent simple linear regressions. WMC = working memory capacity; LBC = list-based clustering index (Stricker et al., 2002); FA = full-attention clustering instructed list. DT = dual-task clustering-instructed list.

CHAPTER IV

DISCUSSION

The present study compared younger and older adults' ability to meaningfully organize study material into semantic clusters on a recall memory task. Spontaneously, older adults semantically clustered study words to the same extent as younger adults. They were further at least as able as younger adults to follow instructions to semantically cluster study words in terms of producing highly clustered recall outputs (RR measure). However, after clustering instructions older adults clustered less successfully (in terms of the number and length of clusters; LBC measure) than younger adults. Consequently, older adults improved their recall less with clustering instructions, resulting in magnified age-related differences in recall compared to when clustering had not been instructed. Thereby, the present study reveals both preservation and limitations of encoding strategies in older adults.

Importantly, presentation format was found to impact the ability to cluster successfully for both age groups. Once semantic clustering use was instructed, participants for whom the study list had been presented as a whole clustered more successfully and had greater recall than those for whom words had been briefly presented individually for the same total study duration. The format conditions did not differ in recall when clustering was not instructed, suggesting that, once instructed, recall was higher in the whole-list conditions due to this format's facilitation of semantic clustering.

Indeed, semantic clustering success fully mediated the format differences in recall on the clustering-instructed lists. Effects for presentation format were uniform across age groups but it is particularly notable that for older adults, who generally benefitted less from the clustering instructions, there was no overall improvement in recall after clustering instructions in the individual-words format. Further, older adults particularly struggled to maintain clustering under DT in the individual-words format, with age-related differences already present in the relative clustering of the recall output (RR) in addition to clustering success (LBC). That is, older adults' ability to use and benefit from semantic clustering was particularly reduced in an individual-words presentation format, especially when resources were limited. Despite such evidence that semantic clustering was more resource-demanding in the individual-words format, there were surprisingly no differences in spontaneous use of clustering between the two formats in either age group. In the following, key implications from the present study will be discussed with a focus on what was learned about cognitive resource demands of semantic clustering, determinants of its spontaneous use, and, importantly, older adults' semantic clustering abilities.

Cognitive Resource Demands of Semantic Clustering

Age-related differences in semantic clustering use are commonly attributed to this strategy's resource demands (Jacobs et al., 2001; Taconnat et al., 2009; Wegesin et al., 2000). To my knowledge, the present study was the first attempt to directly assess evidence for clustering's resource demands by using a dual-task interference procedure (cf., Guttentag, 1984; Kee & Davies, 1990, 1991; Naveh-Benjamin et al., 2005). When tone discriminations had to be performed simultaneously to semantic clustering, participants' responses were slower than when merely reading non-meaningful information on the screen, evidencing that semantic clustering is indeed demanding on general cognitive resources. Further, there was evidence that semantic clustering ability depended on a general cognitive resource, WMC.

Age-related differences in cognitive costs. Given strong age-related declines in general cognitive resources (Park, 2000), it was predicted that older adults would experience greater cognitive costs through semantic clustering. In absolute terms, older adults were slowed more by the addition of semantic clustering to the tone-discrimination task. However, older adults' slower performance at baseline (i.e., non-word reading), in line with typical age-related slowing of cognitive processing (Salthouse, 1996), causes some interpretational ambiguity when comparing the absolute clustering costs because the same increase or decrease in a variable may signify a differential effect at different levels of baseline performance (Loftus, 1985; Somberg & Salthouse, 1982; Wheeler, 2000). If costs were computed relative to baseline performance, no age-related difference was evident. For associative encoding strategies (i.e., imagery and sentence generation), Naveh-Benjamin et al. found higher strategy use costs in older adults, both absolute and relative. In their secondary tracking task, strategy use costs were assessed in terms of accuracy (millimeters away from target) rather than response speed, with accuracy being a measure less likely susceptible to age-related differences (baseline performance not reported). Unfortunately, response speed as a cost measure induces an interpretational ambiguity because of typical age-related slowing present at baseline. Thereby, the higher absolute slowing observed in older adults might be due to higher clustering costs or, simply, their already slower baseline.

Additional findings from the present study suggest that clustering might indeed be particularly demanding for older adults, in line with the absolute cost measure. Subjectively, older adults judged semantic clustering to be more difficult than younger adults. Further, on the FA clustering-instructed list significant positive correlations between WMC and clustering indices emerged in the older adults and across younger and older adults. Indeed, the age-related reductions in WMC mediated the age-related differences in clustering success. Most strikingly, age-related differences emerged in both relative output clustering (RR) and absolute clustering success (LBC) under DT in the individual-words format, suggesting that older adults did not have sufficient resources available even for relative clustering under DT in this format. Thus, even though the RT cost measure results are somewhat inconclusive, there was some support for the hypothesis that semantic clustering was particularly taxing on older adults' available resources. In a future study, a clearer answer might be provided with a secondary task for which baseline performance does not vary with age.

Presentation-format differences in cognitive costs. The whole-list presentation format was predicted to facilitate semantic clustering and this hypothesis was supported by greater clustering success and recall in the whole-list conditions when semantic clustering was used as the (instructed) primary encoding strategy. However, in terms of RT costs, the whole-list format conditions showed greater clustering costs than the individual-words format conditions, both absolute and relative (there were no condition differences in baseline non-word RTs). At the same time, the whole-list format conditions achieved greater clustering and recall under DT. The format conditions received identical instructions for the DT procedure emphasizing the importance of both clustered encoding and tone discriminations, except for references to the presentation format. Thus, it is unlikely that participants were differentially motivated across conditions. Importantly, clustering indices under DT were greater than on the spontaneous list in all conditions, suggesting that participants attempted to follow the clustering instructions. Despite this, participants in the individual-word conditions were not as successful under DT (and even under FA), which suggests that clustering was more difficult and resource-demanding in this format. Importantly, the recall phases were identical for the two formats, hence clustering differences most likely stem from differences at encoding. Participants in the whole-list format may have formed and rehearsed larger clusters during encoding whereas those in the individual-words format may have had difficulty relating back to earlier presented words from the same category during encoding. Thereby, the higher RT clustering costs might rather reflect how much semantic processing was achieved during encoding as opposed to clustering difficulty. This could explain why participants' self-reports of clustering difficulty were not related to RT costs and actually suggested higher clustering difficulty under DT in the individual-words format.

The lower LBC intercepts in the individual-words compared to the whole-list conditions further support the idea that semantic clustering is particularly resource-demanding in the individual-words format. In both formats, clustering success was positively correlated with WMC to a similar extent but at the same level of WMC participants achieved greater clustering success in the whole-list format. That is, low WMC limited successful clustering in the individual-words format but still allowed quite successful clustering in the whole-list format (albeit greater clustering success was possible for those higher in WMC).[9] Further, only in the individual-words format was WMC also related to relative output clustering (RR). Notably, the whole-list conditions performed worse on the WMC tasks but better in terms of clustering and recall (instructed lists). It is unclear whether these randomly assigned conditions really differed in WMC, particularly given no differences in processing speed. Possibly, the individual-words format's match with the presentation mode of the WMC tasks put this condition at an advantage in the WMC tasks. Either way, the better clustering and recall performance of the whole-list conditions cannot be attributed to higher cognitive abilities of participants in these conditions, further suggesting that clustering was indeed less demanding in the whole-list format.

Difficulties in assessing costs of encoding strategies. Siegler and Lemaire (1997) emphasize the importance of controlling strategy use so that benefit and cost aspects of strategies can be accurately estimated independent of the characteristics of the items a strategy is preferably used on and of the characteristics of the individuals who choose to use a certain strategy. But encoding strategies use cannot be as well controlled as strategy use in other areas. For example, Siegler and Lemaire controlled strategy use in

[9] Unsworth and Engle (2007) proposed WMC-related differences in strategic retrieval from long-term memory which may explain part of the WMC relation to clustering success. However, the presentation-format differences suggest that WMC is also related to the ability to group by semantic relatedness during encoding (recall phases were identical for the format conditions), in line with the recent proposal that WMC also influences strategic encoding processes (Unsworth & Spillers, 2010).

arithmetic by providing or removing devices (e.g., calculator, scratch paper). However, encoding strategies are mentally generated rather than dependent on external devices. Compliance with encoding strategy instructions is generally high in both younger and older adults but not perfect (e.g., around 70% for imagery and sentence generation; Dunlosky & Hertzog, 1998; Kuhlmann & Touron, 2012). Little was known about compliance with semantic-clustering instructions prior to this study. In prior studies instructing semantic clustering, participants physically sorted study items printed on index cards into separate stacks (Basden et al., 1993; Guttentag, 1988; Kliegel et al., 2003; Worden & Meggison, 1984). Thereby, grouping was fully enforced but this study setting, which is more akin to the whole-list format, is very different from most memory tasks. In particular, it does not allow examination of *mental* semantic clustering, which was of particular interest for the present study. Other researchers pursuing the same interest of assessing costs of mental encoding strategy use (Kee & Davies, 1990, 1991; Naveh-Benjamin et al., 2005) also solely relied on strategy instructions without further control of encoding activity. Therefore, the present study was the first to employ clustering instructions without allowing physical sorting. All clustering indices suggested high compliance with the semantic clustering instructions, even under DT, but the performance differences complicate comparison across formats.

Determinants of Spontaneous Clustering Use

The role of cognitive costs. Cognitive costs were hypothesized to predict spontaneous clustering use based on theoretical models of strategy choice (Lovett & Anderson, 1996; Shrager & Siegler, 1998; Siegler & Shipley, 1995) as well as empirical findings of correlations between secondary-task costs and spontaneous encoding strategy use in children (Guttentag, 1984; Kee & Davies, 1990, 1991). In the present study, RT clustering costs did not, however, correlate with spontaneous strategy use. Interestingly, self-reported clustering difficulty correlated with spontaneous clustering use such that participants who judged semantic clustering as more difficult were less likely to spontaneously use it. This self-report judgment was made after recall and may thus have been reactive to the achieved clustering although it is notable that these correlations mostly vanished on the FA clustering-instructed list, indicating that participants judged the difficulty somewhat independent from their performance. Difficulty self-reports did not correlate with RT clustering costs. It may be that the chosen operationalization of cognitive clustering costs as tone-discrimination RT interference does not fully capture cost aspects relevant to participants. Further, the disconnection between objective RT costs and subjective difficulty estimates might imply that cost *beliefs* rather than objective costs influence strategy choice. Interestingly, other research found that participants' incorrect beliefs about the speed of a strategy influence their strategy choice (Hertzog et al., 2007; Hines, Touron, & Hertzog, 2009). There were also positive trends for an influence of clustering efficacy beliefs on clustering use, further suggesting a role for metacognitive beliefs in strategy choice (cf., Lachman & Andreoletti, 2006).

Neither difficulty nor efficacy beliefs differed between presentation formats on the spontaneous list. Therefore, a lack of metacognitive appreciation that clustering use is facilitated by the whole-list format could explain the surprising lack of difference in spontaneous clustering use between the formats. Clustering use self-reports also did not differ between the formats on the clustering-instructed lists, suggesting that poor monitoring of achieved clustering may have contributed to this lack of appreciation of the benefits of the whole-list format. Presentation format is an external (not word-inherent) cue to which people's metacognitive judgments are generally insensitive (Koriat, 1997; Touron et al., 2010). Only on the DT clustering-instructed list were there format differences in difficulty judgment; after experiencing a clustering attempt under reduced resources participants in the individual-words format judged clustering as more difficult compared to those in the whole-list format.

Finally, it is important to consider alternative encoding strategies. Participants could use, and did use as indicated in responses on the strategy checklist for the spontaneous list (Appendix F), several other strategies, many of them also effective like the generation of mental images (e.g., Kausler, 1994). Strategy choice models generally put the costs and benefits of one strategy in relation to other available strategies (e.g., Siegler & Lemaire, 1997). The spontaneous list RR index was moderate (30-40%), suggesting that semantic clustering was not the primary strategy used by many participants spontaneously. Words were selected to be typical members of their respective categories but the most typical category exemplars were not included and the words may have had other properties affording alternative strategies. Therefore, if the word material had been more encouraging of semantic clustering, perhaps clustering use differences between the format conditions would have occurred on the spontaneous list.

The role of cognitive resources. Given semantic clustering's demands of maintaining presented words while comparing them to other study words, WMC was predicted to be an important cognitive resource for clustering, especially in the individual-words format where each study word only appeared briefly and different parts of the study list could not be revisited (cf., Jacobs et al., 2001; Wegesin et al., 2000). Consistent with this hypothesis, WMC was strongly and uniquely (beyond processing speed) correlated with older adults' clustering success (LBC; correlation with relative output clustering [RR] in the individual-words format only) on the FA clustering-instructed list. These correlations were only trending in younger adults but remained significant across the age groups. Similarly, WMC-clustering correlations were significant across the age groups on the DT clustering-instructed list (albeit weaker and not significant within each condition but DT has been found to reduce WMC-related differences, e.g. Rosen, 1997). WMC did not, however, correlate with spontaneous clustering. This finding contrasts with Wegesin et al. (2000) who found WMC to be predictive of older adults' spontaneous clustering. The present study was the first to examine the WMC – clustering correlation after instructions to use semantic clustering. The differential findings between the spontaneous and the clustering-instructed lists are particularly informative as they suggest that WMC is an important resource for successful clustering but may not be important for the choice to use clustering as an encoding strategy. The hypothesis that WMC predicts spontaneous clustering posits not only that WMC is an important resource for clustering but also that clustering thereby is particularly effortful for people with low WMC and, further, that such cognitive effort or costs influence spontaneous strategy choice. Despite its relation to clustering success, WMC was not related to objective RT clustering costs or subjective clustering difficulty.

That is, even though at least older adults with high WMC (and tentatively younger adults) were better able to semantically cluster study words this did not translate into them experiencing clustering as less difficult.

Nonetheless, Wegesin et al. (2000) found WMC to be positively correlated with older adults' spontaneous clustering. Further, Taconnat et al. (2009) as well as Jacobs et al. (2001) Wegesin et al. found positive correlations between executive functioning measures and spontaneous clustering use in both younger and older adults. These measures are strongly correlated with WMC measures and it has been proposed that both commonly measure basic attention control abilities (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Notably, there is good evidence for distinctive components of executive functioning (Miyake et al., 2000). Thus, whereas WMC appears to be a necessary resource for the execution of semantic clustering, it may be that strategy choice processes are influenced by a more specific aspect of executive functioning. An important first step in spontaneous strategy use is recognizing that a strategy can be used on a task (Lemaire, 2010). For semantic clustering, this involves recognizing that the material can be reorganized from its original presentation form, requiring what Taconnat et al. deemed "cognitive flexibility." Thereby, mental set shifting, one of the executive functioning components identified by Miyake et al., might play a particular role in recognizing that words can be reorganized by semantic category. Interestingly, Bouazzaoui et al. (2010) found executive functioning measures to predict reported daily-life use of encoding strategies and to mediate age-related differences therein. Future research involving a battery of executive functioning tasks in addition to WMC measures is needed to disentangle their contributions to spontaneous use of semantic clustering and other encoding strategies.

Alternatively, it is possible that in the prior studies finding WMC or executive functioning to correlate with spontaneous clustering use, the word material or other aspects of the task made participants more likely to use semantic clustering as the primary encoding strategy, making it more alike to the clustering-instructed lists of the present study. In that case, the correlation would reflect a relation of WMC or executive functioning to strategy execution ability rather than strategy choice. It is notable that younger adults spontaneously produced very highly clustered outputs in Taconnat et al. (2009; mean ARC = .70) suggesting much higher clustering inclination than in the present study (mean .27, see Appendix D) although rates in their older adult sample were comparable to the present study (mean ARC = .39 vs. 36).

Older Adults' Semantic Clustering Ability

Production. A review of past research suggested that age-related differences in spontaneous semantic clustering use primarily occurred when the study list was presented with individual words as opposed to the whole list. Of course, these studies also differ in other potentially relevant methodological aspects like word material, list composition (number of categories and words per category), and presentation time. The current study was the first to systematically examine presentation format keeping all else constant. But presentation format affected neither older nor younger adults' spontaneous clustering, with no evidence for an age-related production deficiency in either format. Thus, there was no evidence that presentation format moderates the occurrence of age-related production deficiences for semantic clustering.

Still, the findings from the clustering-instructed lists do suggest a particular vulnerability of older adults' clustering in the individual-words format. Similarly, Park et al. (1989) found older adults' spontaneous clustering use (measured by ARC, similar to RR) in an individual-words format to be particularly reduced under divided attention at encoding. The present findings nicely complement this study because the finding of a similar DT-induced age-related clustering difference after instructions suggests that DT does not merely render older adults less inclined to cluster but rather affects their ability to cluster. Further, the present study demonstrates that this finding is specific to the individual-words format. Quite contrary, older adults' recall of the words was more clustered than in younger adults in the whole-list format, even under DT. That is, a simple change in presentation format can greatly aid older adults' semantic clustering.

Consequently, even though no presentation-format effects on spontaneous clustering were found in the present study, these findings suggest that older adults are particularly sensitive to list presentation format and that age-related differences are most likely to occur in the individual-words format. Future research is needed to determine the particular conditions under which a produdction deficiency occurs in this format. List composition such as the number of categories as well as the typicality of the exemplars might be relevant factors (Howard et al., 1981; Witte et al., 1993). For example, fewer words per category placed further apart within the study list may particularly reduce older adults' clustering use in the individual-words format.

As already discussed, spontaneous clustering in the present study was rather low. Age-related differences in clustering might be more likely when there is high motivation to cluster. In terms of absolute clustering success (LBC measure), younger adults outperformed older adults when instructed to use semantic clustering. Notably, the choice of clustering measure may also play a role as in the present study the RR and LBC measure differed with regards to age-group effects on the clustering-instructed lists. In particular, no age-related reduction (with the exception of the DT clustering-instructed list in the individual-words format) occurred in RR, with older adults in the whole-list condition even scoring higher on this relative clustering index. For LBC, which is an absolute clustering success measure rewarding greater clustered recall, however, age-related differences occurred on the clustering-instructed lists in both formats. Some studies have found age-related production deficiencies in relative output clustering measures like RR and ARC (Amrhein et al., 1999; Schneider & Uhl, 1990; Taconnat et al., 2009). But Jacobs et al. (2001; see also Wegesin et al., 2000) used LBC and do not report if age-related differences held in a relative measure. Notably, though, in the present study age-related differences on the spontaneous list were absent in both RR and LBC.

Finally, occurrence of age-related clustering production deficiencies may also depend on metacognitive factors. The present older adult sample expressed similar beliefs about an internal locus of memory control and clustering efficacy. Other research has found older adults to be less likely to believe that memory can be strategically controlled (Lachman, 2006) although like in the present study control beliefs are often not related to actual strategy use (Blatt-Eisengart & Lachman, 2004; Hertzog et al., 2010; but see Hertzog et al., 1998; Lachman & Andreoletti, 2006). As is typical, however, the current older adult sample reported less frequent use of internal encoding strategy use in daily life and also did not appear overly strategic in strategy reports for the spontaneous list (compare Appendix F). Generally, older adults showed good metacognitive monitoring (cf., Hertzog & Hultsch, 2000); their self-reports of clustering use reflected the age-related reduction in clustering success and were positively correlated with clustering indices to the same extent as in younger adults. But, like younger adults, older adults' metacognitive judgments did not reflect format differences (except for difficulty judgments under DT), even though format effects were even somewhat stronger in older adults (for the RR measure). Thereby, older adults did not spontaneously recognize a memory task setting in which they could have very much benefitted from semantic clustering. Thus, metacognitive awareness might be crucial for task affordability effects on spontaneous strategy production.

Utilization. Generally, semantic clustering is regarded as a beneficial strategy for both younger and older adults (Kausler, 1994) but recently Taconnat et al. (2009) questioned that older adults can improve their memory through clustering as they found no (and sometimes even negative) correlations between clustering and recall in older adults (in an individual-words format). Many studies do not examine the clustering – recall relationship separately by age group; the few that do found positive correlations in the older adults that were comparable to those for younger adults (Hess et al., 2003; Lachman, 2006; Witte et al., 1993) or even higher (Schneider & Uhl, 1990). In the present study, there were generally significant positive cluster - recall correlations in older adults that were comparable to those observed in younger adults. One exception, are the nonsignificant RR and recall correlations in the older adult whole-list conditions on the clustering-instructed lists. There were a few older adults in this condition that only recalled very few items but in a highly clustered manner (e.g., only items from one category). This is an issue of the relative output clustering measures like RR; for LBC indexing clustering success significant positive correlations emerged in all conditions and on all lists.

Most importantly, the examination of clustering instruction effects on recall in the present study allows a better examination of clustering benefits unconfounded with the individual characteristics of spontaneous clustering users (cf., Siegler & Lemaire, 1997). In the individual-words format, older adults' average recall performance did not increase with clustering instructions whereas for older adults in the whole-list format recall significantly increased. However, compared to younger adults, the recall increase in this format in older adults (9%) was reduced compared to younger adults (17%). Thus, there was some support for Taconnat et al. with reduced clustering benefits (if any) in older adults. But it would be wrong to conclude that older adults generally cannot benefit from semantic clustering. For one, the positive clustering - recall correlation suggests that some older adults can benefit from clustering in the individual-words format. Further, many older adults benefitted in the whole-list format. These findings are in line with the life-span perspective which recognizes plasticity of older adults' cognitive performance but also limitations of this plasticity compared to younger adults and children (cf., Baltes & Baltes, 1993; Brehmer, Li, Müller, von Oertzen, & Lindenberger, 2007).

The reduced clustering benefits in older adults resulted in larger age-related recall differences after strategy instructions. A similar magnification of age differences has been reported after instructions in the method of loci (Baltes & Kliegl, 1992) and has been related to this mnemonic's cognitive resource demands (Verhaeghen & Marcoen, 1996). Comparably, WMC mediated age-related differences in clustering success on the instructed lists in the present study. For other encoding strategies (e.g., imagery and sentence generation), however, memory benefits in older adults were comparable to those seen in younger adults (e.g., Dunlosky & Hertzog, 2001; Kuhlmann & Touron, 2012). Thus, semantic clustering is a particularly challenging encoding strategy for older adults, especially in the individual-words presentation format, resulting in a utilization deficiency.

Conclusion

In summary, the present study suggests that a simple change in presentation format to presenting the whole list at once for study rather than words individually enables more successful semantic clustering in younger and older adults, resulting in higher recall performance. Importantly, older adults were only able to overall improve their recall with instructions to semantically cluster when studying in a whole-list but not in an individual-words format. Successful clustering depended on WMC, especially in the individual-words format where lower levels of WMC resulted in lower levels of clustering than in the whole-list format. However, spontaneous clustering did not differ between presentation formats for either age group. Metacognitive judgments did not suggest awareness of presentation-format benefits. Thus, simple changes in list presentation format can greatly influence older (and younger) adults' recall performance by facilitating strategy use but do not necessarily result in these benefits unless strategy use is encouraged.

REFERENCES

Amrhein, P. C., Bond, J. K., & Hamilton, D. A. (1999). Locus of control and the age difference in free recall from episodic memory. *The Journal of General Psychology*, *126*, 149–164. doi:10.1080/00221309909595358

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24*, 82-92. doi:10.1037/a0014078

Baltes, P. B., & Baltes, M. M. (1993). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 1-34). New York, NY, US: Cambridge Univ Press.

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology, 28*, 121-125. doi:10.1037/0012-1649.28.1.121

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Basden, B. H., Basden, D. R., & Bartlett, K. (1993). Memory and organization in elderly subjects. *Experimental Aging Research, 19*, 29–38. doi:10.1080/03610739308253921

Blatt-Eisengart, I., & Lachman, M. E. (2004). Attributions for memory performance in adulthood: Age differences and mediation effects. *Aging, Neuropsychology, and Cognition*, *11*, 68-79. doi:10.1076/anec.11.1.68.29364

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journal of Gerontology: Psychological Sciences, 60*, 223-233. doi:10.1093/geronb/60.5.P223

Bottiroli, S., Dunlosky, J., Guerini, K., Cavallini, E., & Hertzog, C. (2010). Does task affordance moderate age-related deficits in strategy production? *Aging, Neuropsychology, and Cognition, 17,* 591-602. doi:10.1080/13825585.2010.481356

Bouazzaoui, B., Isingrini, M., Fay, S., Angel, L., Vanneste, S., Clarys, D., & Taconnat, L. (2010). Aging and self-reported internal and external memory strategy uses: The role of executive functioning. *Acta Psychologica, 135*(1), 59-66. doi:10.1016/j.actpsy.2010.05.007

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology, 49*, 229-240. doi:10.1080/00221309.1953.9710088

Brehmer, Y., Li, S.-C., Müller, V., von Oertzen, T., & Lindenberger, U. (2007). Memory plasticity across the life span: Uncovering children's latent potential. *Developmental Psychology, 43*(2), 465-478. doi:10.1037/0012-1649.43.2.465

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*(3), 249-257. doi:10.1037//0882-7974.3.3.249

Cavallini, E., Pagnin, A., & Vecchi, T. (2003). Aging and everyday memory: The beneficial effect of memory training. *Archives of Gerontology and Geriatrics, 37*, 241-257. doi: 10.1016/S0167-4943(03)00063-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409-421). Oxford, England: North Holland.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes (Advances in the study of communication and affect, Vol. 8)* (pp. 191-211). New York, NY, USA: Plenum Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. doi: doi:10.1016/S0022-5371(72)80001-X

Derwinger, A., Stigsdotter Neely, A., & Bäckman, L. (2005). Design your own memory strategies! Self-generated strategy training versus mnemonic training in old age: An 8-month follow-up. *Neuropsychological Rehabilitation, 15*, 37-54. doi:10.1080/09602010343000336

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology, 38*, 682-688.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597-607. doi:10.1037/0882-7974.13.4.597

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29*, 247-253. doi:10.3758/BF03194918

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*, 3–84.

Fernandes, M. A., & Grady, C. (2008). Age differences in susceptibility to memory interference during recall of categorizable but not unrelated word lists. *Experimental Aging Research, 34*, 297-322. doi:10.1080/03610730802273860

Frankel, F., & Cole, M. (1971). Measures of category clustering in free recall. *Psychological Bulletin, 76*, 39-44. doi:10.1037/h0031256

Frender, R., & Doubilet, P. (1974). More on measures of category clustering in free recall-although probably not the last word. *Psychological Bulletin, 81*, 64-66. doi:10.1037/h0035493

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology, 106*, 92-106. doi:10.1016/0022-0965(84)90060-2

Guttentag, R. E. (1988). Processing relational and item-specific information: Effects of aging and division of attention. *Canadian Journal of Psychology, 42*, 414-423. doi:10.1037/h0084205

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging, 5*, 215-227. doi:10.1037/0882-7974.5.2.215

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd edition ed., pp. 417-466). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment Effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest, 9*, 1-65. doi: 10.1111/j.1539-6053.2009.01034.x

Hertzog, C., McGuire, C. L., Horhota, M., & Jopp, D. (2010). Does believing in 'use it or lose it' relate to self-rated memory control, strategy use, and recall? *The*

International Journal of Aging and Human Development, 70, 61-87. doi:10.2190/AG.70.1.c

Hertzog, C., McGuire, C. L., & Lineweaver, T. T. (1998). Aging, attributions, perceived control, and strategy use in a free recall task. *Aging, Neuropsychology, and Cognition, 5,* 85-106. doi:10.1076/anec.5.2.85.601

Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time-monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychology and Aging, 22*, 607-624. doi:10.1037/0882-7974.22.3.607

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *Journal of Gerontology: Psychological Sciences*, *58B*, 3-11. doi:10.1093/geronb/58.1.P3

Hess, T. M., Germain, C. M., Swaim, E. L., & Osowski, N. L. (2009). Aging and selective engagement: The moderating impact of motivation on older adults' resource utilization. *Journal of Gerontology: Psychological Sciences, 64B*, 447-456. doi:10.1093/geronb/gbp020.

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*, 462-475. doi:10.1037/a0014417

Howard, D. V., McAndrews, M. P., & Lasaga, M. I. (1981). Semantic priming of lexical decisions in young and old adults. *Journal of Gerontology, 36*, 707-714. doi:10.1093/geronj/36.6.707

Hulicka, I. M., & Grossman, J. L. (1967). Age-group comparisons for the use of mediators in paired-associate learning. *Journal of Gerontology*, *22*, 46-51.

Hultsch, D. F. (1971). Adult age differences in free classification and free recall. *Developmental Psychology, 4*, 338-342. doi:10.1037/h0030990

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514. doi:10.1016/S0022-5371(81)90138-9

Jacobs, D. M., Rakitin, B. C., Zubin, N. R., Ventura, P. R., & Stern, Y. (2001). Cognitive correlates of mnemonics usage and verbal recall memory in old age. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 14*, 15-22.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217. doi:10.1037/0096-3445.133.2.189

Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA, USA: Academic Press.

Kee, D. W., & Davies, L. (1990). Mental effort and elaboration: Effects of accessibility and instruction. *Journal of Experimental Child Psychology, 49*, 264-274. doi:10.1016/0022-0965(90)90058-G

Kee, D. W., & Davies, L. (1991). Mental effort and elaboration: A developmental analysis of accessibility effects. *Journal of Experimental Child Psychology, 52*, 1-10. doi:10.1016/0022-0965(91)90002-A

Kliegel, M., Altgassen, M., Martin, M., & Kruse, A. (2003). Lernen im Alter: Die Bedeutung der selbstständigen Strukturierung des Lernmaterials [Learning in old age: The importance of self-initiated structuring]. *Zeitschrift für Gerontologie und Geriatrie, 36*, 421-428.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25*, 247-256. doi:10.1037/0012-1649.25.2.247

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, 894-904. doi:10.1037//0012-1649.26.6.894

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370. doi:10.1037//0096-3445.126.4.349

Kuhlmann, B. G., & Touron, D. R. (2012). Mediator-based encoding strategies in source monitoring in young and older adults. *Journal of Experimental Pychology: Learning, Memory, and Cognition, 38*, 1352-1364. doi:10.1037/a0027863

Lachman, M. E. (2006). Perceived control over aging-related declines: Adaptive beliefs and behaviors. *Current Directions in Psychological Science, 15*, 282-286. doi:10.1111/j.1467-8721.2006.00453.x

Lachman, M. E., & Andreoletti, C. (2006). Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults. *Journal of Gerontology: Psychological Sciences, 61B*, 88-94. doi:10.1093/geronb/61.2.P88

Lemaire, P. (2010). Cognitive strategy variations during aging. *Current Directions in Psychological Science, 19*, 363-369. doi:10.1177/0963721410390354

Levine, D. W., & Dunlap, W. P. (1982). Power of the F test with skewed data: Should one transform or not? *Psychological Bulletin, 92*, 272-280. doi:10.1037//0033-2909.92.1.272

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 397-406. doi:10.1037/0278-7393.11.2.397

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology, 217*, 168-217. doi:10.1006/cogp.1996.0016

Mason, S. E., & Smith, A. D. (1977). Imagery in the aged. *Experimental Aging Research, 3*, 17-32. doi:10.1080/03610737708257085

Mata, R., von Helversen, B., & Rieskamp, J. (2010). Learning to choose: Cognitive aging and strategy selection learning in decision making. *Psychology and Aging, 25*, 299-309. doi:10.1037/a0018923

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology, 24*, 222-243. doi:10.1037/a0017619

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100. doi:10.1006/cogp.1999.0734

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125. doi:10.1037//1082-989X.7.1.105

Mungas, D., Ehlers, C. L., & Blunden, D. (1991). Age differences in recall and information processing verbal and spatial learning. *Canadian Journal on Aging/La revuew canadienne du viellissement, 10*, 320-332. doi:10.1017/S0714980800011351

Murphy, M. D. (1979). Measurement of category clustering in free recall. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 51-83). New York, NY, USA: Academic Press.

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging, 22*(1), 202-208. doi:10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 520-537. doi:10.1037/0278-7393.31.3.520

Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. C. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 3-19). Philadelphia, PA, US: Psychology Press.

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1185-1191. doi:10.1037/0278-7393.15.6.1185

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891. doi:10.3758/BRM.40.3.879

Rankin, J. L., Karol, R., & Tuten, C. (1984). Strategy use, recall, and recall organization in young, middle-aged, and elderly adults. *Experimental Aging Research, 10*, 193-196. doi:10.1080/03610738408258463

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*, 45-48. doi:10.1037/h0031355

Rosen, V. M. E., Randall W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211-227. doi:10.1037/0096-3445.126.3.211

Rowe, E. J., & Schnore, M. M. (1971). Item concreteness and reported strategies in paired-associate learning as a function of age. *Journal of Gerontology, 26*, 470-475. doi:10.1093/geronj/26.4.470

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428. doi:10.1037/0033-295X.103.3.403

Schneider, W., & Uhl, C. (1990). Metagedächtnis, Strategienutzung und Gedächtnisleistung: Vergleichende Analysen bei Kindern, jüngeren Erwachsenen und alten Menschen [Metamemory, strategies, and memory performance: Comparative analysis among children, young adults, and older adults]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *22*, 22-41.

Shrager, J., & Siegler, R. S. (1998). SCADS : A model of children’s strategy choices and strategy discoveries. *Psychological Science, 9*, 405-410. doi:10.1111/1467-9280.00076

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY, US: Oxford University Press.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*, 71-92. doi:10.1037/0096-3445.126.1.71

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Somberg, B., & Salthouse, T. A. (1982). Divided attention abilities in young and old adults. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 651–663. doi:10.1037/0096-1523.8.5.651

Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test – Second Edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, *8*, 425-435. doi:10.1017/S1355617702813224

Taconnat, L., Raz, N., Toczé, C., Bouazzaoui, B., Sauzéon, H., Fay, S., & Isingrini, M. (2009). Ageing and organisation strategies in free recall: The role of cognitive flexibility. *European Journal of Cognitive Psychology*, *21*, 347-365. doi:10.1080/09541440802296413

Tournier, I., & Postal, V. (2011). Strategy selection and aging: Impact of item concreteness in paired-associate task. *Aging, Neuropsychology, and Cognition, 18*, 195-213. doi: 10.1080/13825585.2010.525623

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19,* 452-466. doi:10.1037/0882-7974.19.3.452

Touron, D. R., & Hertzog, C. (2004b). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition, 32*, 298-310. doi:10.3758/BF03196860

Touron, D. R., Hertzog, C., & Speagle, J. Z. (2010). Subjective learning discounts test type: Evidence from an associative learning and transfer task. *Experimental Psychology*, *57*, 327-337. doi: 10.1027/1618-3169/a000039

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132. doi:10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505. doi:10.3758/BF03192720

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review, 17*, 200-205. doi:10.3758/PBR.17.2.200

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50*, 289-335. doi: 10.1016/j.jml.2003.10.003

Verhaeghen, P., & Marcoen, A. (1994). Production deficiency hypothesis revisited: Adult age differences in strategy use as a function of processing resources. *Aging, Neuropsychology, and Cognition, 1*, 323-338. doi:10.1080/13825589408256585

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging, 11*, 164-178. doi:10.1037/0882-7974.11.1.164

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7*, 242-251. doi:10.1037/0882-7974.8.3.338

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York, NY, US: Psychological Corporation.

Wegesin, D. J., Jacobs, D. M., Zubin, N. R., Ventura, P. R., & Stern, Y. (2000). Source memory and encoding strategy in normal aging. *Journal of Clinical and Experimental Neuropsychology, 22*, 455-464. doi: 10.1076/1380-3395(200008)22:4;1-0;FT455

West, R. L., Dark-Freudeman, A., & Bagwell, D. K. (2009). Goals-feedback conditions and episodic memory: Mechanisms for memory gains in older and younger adults. *Memory, 17*, 233-244. doi:10.1080/09658210802236407

West, R. L., & Thorn, R. M. (2001). Goal-setting, self-efficacy, and memory performance in older and younger adults. *Experimental Aging Research, 27*, 41-65. doi:10.1080/036107301750046133

Wheeler, M. A. (2000). A comparison of forgetting rates in older and younger adults. *Aging, Neuropsychology, and Cognition, 7*, 179-193. doi:10.1076/1382-5585(200009)7:3;1-Q;FT179

Witte, K. L., Freund, J. S., & Brown-Whistler, S. (1993). Adult age differences in free recall and category clustering. *Experimental Aging Research*, *19*, 15-28. doi:10.1080/03610739308253920

Worden, P. E., & Meggison, D. L. (1984). Aging and the category-recall relationship. *Journal of Gerontology, 39*, 322-324. doi:10.1093/geronj/39.3.322

Zivian, M., & Darjes, R. W. (1983). Free recall by in-school and out-of-school adults: Performance and metamemory. *Developmental Psychology, 19*, 513-520. doi:10.1037/0012-1649.19.4.513

APPENDIX A

CATEGORIZABLE LISTS USED IN THE EXPERIMENT

List A	List B	List C
Animals	*Clothes*	*Birds*
horse	socks	eagle
lion	hat	robin
bear	jacket	bluejay
tiger	sweater	cardinal
elephant	skirt	hawk
Vegetables	*Musical Instruments*	*Fruits*
carrot	drum	banana
broccoli	guitar	grape
peas	flute	peach
corn	piano	strawberry
onion	cello	plum
Sports	*Vehicles*	*Metals*
soccer	bus	steel
baseball	truck	iron
tennis	plane	silver
hockey	bike	copper
golf	boat	gold
Body parts	*Flowers*	*Furniture*
foot	daisy	table
finger	tulip	couch
head	lily	bed
nose	carnation	desk
ear	daffodil	dresser
M named = .49 (.04)	*M* named = .48 (.04)	*M* named = .49 (.03)
M syllables = 1.65 (.15)	*M* syllables = 1.65 (.17)	*M* syllables = 1.70 (.16)
M letters = 5.2 (.34)	*M* letters = 5.1 (.32)	*M* letters = 5.4 (.37)

Note. Standard errors in parentheses. Italicized category labels were not presented for study. *M* named = mean proportion with which exemplars were named by norming sample in response to category label (Van Overschelde et al., 2004).

APPENDIX B

METACOGNITIVE QUESTIONS AFTER LIST RECALL

Note: Question 1 was only asked after the first (spontaneous) list. Estimation of hypothetical difficulty in questions 4 and 5 was only asked after the first list; after the second and third list participants were asked to directly indicate the actual difficulty since they had been instructed to use clustering. Horizontal lines indicate a new computer screen.

1) We are interested in the types of strategies people use to learn a list like the one you were asked to learn today. Below are descriptions of strategies people sometimes report using.

- ☐ Mentally picturing the presented words. For example, picturing a carpenter with a hammer when studying the word 'hammer.'
- ☐ Making sentences or stories with the presented words. For example, when studying 'nail' and 'oak' one might have said "Thick nails are needed to hold the heavy oak-wood frame."
- ☐ Grouping words from the same category. For example, 'hammer' and 'nail' are tools while 'oak' and 'pine' are trees.
- ☐ Groupings words by other criteria, for example words that start with the same letter or words that rhyme.
- ☐ Mentally repeating the words, either by themselves (like 'nail, nail, nail') or in groups (like 'hammer, nail, oak, hammer, nail, oak').
- ☐ Using some other strategy.

If you used a particular strategy while studying the words on the last list, click on the box next to it with the mouse (an X will appear). You may select more than one strategy. Click on a box for a second time to deselect. Click the SUBMIT button below when you have selected all strategies you used. If you did not use any particular strategy, click the button below without having selected a strategy.

The words on the list you just studied could be grouped by category. For example, the words "hammer" and "nail" belong to the category TOOLS and the words "oak" and "pine" belong to the category TREES. These particular words were not on the list you just studied but words on the list you just studied could be grouped in a similar way.

Next, we have a few questions about your use of and your beliefs about this strategy of grouping words by category. Importantly, one may group words by category during STUDY of the words or when being TESTED on the words or during both study and test. Some of the following questions will ask about grouping during STUDY and others will ask about grouping during TESTING. Many questions will be worded similarly but differ with regard to STUDY or TESTING so make sure you read each new question carefully before answering

---

2) As you were STUDYING the last list of words, how many (if any) of the words did you group by category?

0% - 10% - 20% - 30% - 40% - 50% - 60% - 70% - 80% - 90% - 100%

---

3) As you were being TESTED on the last list of words, how many (if any) of the words did you group by category?

0% - 10% - 20% - 30% - 40% - 50% - 60% - 70% - 80% - 90% - 100%

---

4) While STUDYING the last list of words, how difficult did you find grouping the words on the list by category? If you did not group the words during study, estimate how difficult you think grouping the words by category would be for a list like the one you just studied.

not at all difficult	slightly difficult	somewhat difficult	quite difficult	very difficult

---

5) While being TESTED on the last list of words, how difficult did you find grouping the words on the list by category? If you did not group the words during testing, estimate how difficult you think grouping words by category would be for a list like the one you just studied.

not at all difficult	slightly difficult	somewhat difficult	quite difficult	very difficult

---

6) To what extent do you think grouping words by category during STUDY could improve YOUR MEMORY for a list like the one you just studied?

not at all difficult	slightly difficult	somewhat difficult	quite difficult	very difficult

---

7) To what extent do you think grouping words by category during STUDY could improve MEMORY FOR OTHER PEOPLE YOUR AGE?

not at all	slightly	somewhat	quite	very much

---

8) To what extent do you think grouping words by category during TESTING could improve YOUR MEMORY for a list like the one you just studied?

not at all	slightly	somewhat	quite	very much

---

9) To what extent do you think grouping words by category during TESTING could improve MEMORY FOR OTHER PEOPLE YOUR AGE?

not at all	slightly	somewhat	quite	very much

APPENDIX C

FULL CORRELATION MATRICES

Table C1

Full Correlation Matrix for Younger Adults in the Individual-Words Format

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) SP RR													
2) FA RR	.236^{+}												
3) DT RR	.246*	.183											
4) SP LBC	.956*	.296*	.257*										
5) FA LBC	.336*	.928*	.259*	.392*									
6) DT LBC	.319*	.318*	.735*	.349*	.430*								
7) SP Self-Reported Cluster Use	.468*	.054	.260*	.517*	.102	.265*							
8) FA Self-Reported Cluster Use	.107	.283*	.099	.070	.295*	.237^{+}	.310*						
9) DT Self-Reported Cluster Use	.200	.168	.268*	.173	.204^{+}	.506*	.290*	.614*					
10) SP Recall	.543*	.014	.201	.625*	.159	.342*	.546*	-.004	.143				
11) FA Recall	.296*	.373*	.237^{+}	.335*	.662*	.434*	.223^{+}	.241^{+}	.211^{+}	.420*			
12) DT Recall	.262*	.082	.383*	.249*	.215^{+}	.792*	.215^{+}	.176	.508*	.373*	.421*		
13) Education Years	.092	-.088	.031	.034	-.077	.097	-.051	.130	.208^{+}	-.033	-.025	.175	
14) Vocabulary	.062	.044	.011	.028	.106	.130	.166	.243*	.151	.065	.225^{+}	.217^{+}	-.019
15) Speed	.061	.015	.293*	.104	.079	.070	.147	.072	.108	.233^{+}	.166	-.008	-.088
16) WMC	.151	.212^{+}	.185	.160	.211^{+}	.228^{+}	.155	-.014	.180	.181	.136	.298*	.047
17) RT Clustering Costs	-.109	.264*	.144	-.126	.234^{+}	.167	-.064	.074	.124	-.196	.076	.195	-.005
18) SP Difficulty	-.236^{+}	.167	-.018	-.247*	.105	-.041	-.396*	-.243*	-.161	-.324*	-.093	-.118	-.128
19) FA Difficulty	.149	-.025	-.033	.164	-.049	-.235^{+}	-.153	-.337*	-.120	.061	-.170	-.272*	.101
20) DT Difficulty	.023	.131	-.170	.053	.048	-.418*	-.023	-.170	-.424*	-.065	-.105	-.552*	-.055
21) SP Efficacy Belief	.233^{+}	.016	.147	.231^{+}	.035	.081	.550*	.133	.058	.124	.070	-.014	-.118
22) FA Efficacy Belief	.034	.047	.059	-.016	.070	.052	.279*	.272*	.188	-.147	.143	-.058	-.131
23) DT Efficacy Belief	.097	.059	.278*	.077	.076	.189	.272*	.116	.163	-.019	.056	-.003	-.127
24) MIA Achievement	-.114	.112	.036	-.133	.119	.077	.026	.104	.067	-.120	.091	.027	-.010
25) MIA Locus	.029	.171	.083	-.004	.137	.147	.103	.153	.109	-.012	.040	.110	.057
26) MIA Internal Strategies	-.200	.118	.107	-.226^{+}	.090	.016	-.173	-.152	-.015	-.237^{+}	-.041	-.028	-.050
27) MIA External Strategies	.023	.019	-.141	-.016	.026	-.080	-.002	.016	.027	-.046	.033	.128	.127

Table C1 (continued)

Measure	14	15	16	17	18	19	20	21	22	23	24	25	26
1) SP RR													
2) FA RR													
3) DT RR													
4) SP LBC													
5) FA LBC													
6) DT LBC													
7) SP Self-Reported Cluster Use													
8) FA Self-Reported Cluster Use													
9) DT Self-Reported Cluster Use													
10) SP Recall													
11) FA Recall													
12) DT Recall													
13) Education Years													
14) Vocabulary													
15) Speed	.126												
16) WMC	.228^{+}	.251*											
17) RT Clustering Costs	.182	-.134	.120										
18) SP Difficulty	-.078	.091	.138	.195									
19) FA Difficulty	-.002	.162	.170	.009	.222^{+}								
20) DT Difficulty	-.242^{+}	.060	.124	-.026	.216^{+}	.427*							
21) SP Efficacy Belief	.116	-.043	.120	-.080	-.404*	-.138	.077						
22) FA Efficacy Belief	.130	-.069	-.014	.031	-.216^{+}	-.262*	-.066	.660*					
23) DT Efficacy Belief	-.200	-.075	-.047	-.046	-.092	-.129	-.104	.503*	.639*				
24) MIA Achievement	.109	-.164	-.157	-.055	-.041	-.190	-.157	.217^{+}	.305*	.159			
25) MIA Locus	-.024	-.039	.085	-.169	.011	-.198	-.123	.189	.156	.130	.450*		
26) MIA Internal Strategies	-.200	-.094	.008	-.024	.246^{+}	.014	.097	.012	.094	.197	.502*	.320*	
27) MIA External Strategies	.021	-.112	-.062	.009	-.132	-.031	.067	.111	.187	.128	.036	.092	.276*

Note. Spearman *rho* used for self-reported cluster use, difficulty and efficacy belief reports which were made on ordinal Likert scales. Pearson r is used for all other variables. SP = spontaneous list; FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; MIA = Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983).

* $p \leq .05$; $^{+}$ $.05 < p \leq .10$.

Table C2

Full Correlation Matrix for Younger Adults in the Whole-List Format

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) SP RR													
2) FA RR	.305*												
3) DT RR	.180	.578*											
4) SP LBC	.969*	.328*	.229$^+$										
5) FA LBC	.287*	.939*	.571*	.315*									
6) DT LBC	.199	.482*	.890*	.267*	.527*								
7) SP Self-Reported Cluster Use	.724*	.252*	.298*	.690*	.228$^+$	.241$^+$							
8) FA Self-Reported Cluster Use	.069	.486*	.311*	.100	.588*	.322*	.175						
9) DT Self-Reported Cluster Use	.063	.354*	.547*	.094	.383*	.584*	.189	.603*					
10) SP Recall	.607*	.152	.102	.610*	.191	.141	.598*	.090	-.032				
11) FA Recall	.119	.426*	.311*	.153	.649*	.356*	.162	.619*	.320*	.237$^+$			
12) DT Recall	.153	.422*	.592*	.244*	.466*	.741*	.198	.389*	.593*	.223$^+$	.486*		
13) Education Years	.089	.163	-.044	.033	.198	.014	.117	.018	.027	.059	.062	.065	
14) Vocabulary	-.186	-.015	.000	-.136	.019	.006	-.193	-.064	-.088	.269*	.227$^+$	.135	.065
15) Speed	.151	.018	.233$^+$	.174	.042	.241$^+$	.172	.031	.042	.362*	.102	.188	.177
16) WMC	.045	.165	.079	.089	.125	.020	.014	.319*	.312*	.297*	.155	.221$^+$	-.008
17) RT Clustering Costs	.103	.323*	.463*	.128	.323*	.414*	.124	.100	.201	.091	.216$^+$	.417*	-.132
18) SP Difficulty	-.373*	-.187	-.205$^+$	-.377*	-.161	-.165	-.338*	-.258*	-.193	-.236$^+$	-.034	-.099	.130
19) FA Difficulty	.121	-.128	-.097	.087	-.200	-.070	.189	-.346*	-.114	.252*	-.301*	-.014	.082
20) DT Difficulty	.129	.037	-.245*	.107	.060	-.202	.029	-.123	-.344*	.261*	.104	-.192	-.041
21) SP Efficacy Belief	.389*	.098	.335*	.422*	.094	.283*	.414*	.227$^+$	.300*	.269*	.118	.260*	-.217$^+$
22) FA Efficacy Belief	-.122	.226$^+$	.192	-.077	.230$^+$	.197	-.153	.378*	.310*	-.214$^+$	.306*	.223$^+$	-.189
23) DT Efficacy Belief	-.049	.077	.370*	.011	.130	.367*	.019	.222$^+$	.485*	-.147	.142	.423*	-.109
24) MIA Achievement	-.010	.044	-.140	.007	.109	-.033	.021	.040	.029	.096	.116	.125	.125
25) MIA Locus	-.073	-.114	-.295*	-.108	-.048	-.192	-.158	-.056	-.137	-.137	.112	-.013	.090
26) MIA Internal Strategies	-.033	-.071	-.198	-.062	-.074	-.243*	-.079	-.035	-.172	-.068	-.087	-.211$^+$	.033
27) MIA External Strategies	-.017	.107	-.032	-.056	.041	-.170	-.215$^+$	.056	.076	-.055	-.068	-.091	.051

Table C2 (continued)

Measure	14	15	16	17	18	19	20	21	22	23	24	25	26
1) SP RR													
2) FA RR													
3) DT RR													
4) SP LBC													
5) FA LBC													
6) DT LBC													
7) SP Self-Reported Cluster Use													
8) FA Self-Reported Cluster Use													
9) DT Self-Reported Cluster Use													
10) SP Recall													
11) FA Recall													
12) DT Recall													
13) Education Years													
14) Vocabulary													
15) Speed	.196												
16) WMC	.447*	.237$^+$											
17) RT Clustering Costs	.006	.081	.135										
18) SP Difficulty	.084	-.171	-.171	-.116									
19) FA Difficulty	.171	-.129	-.008	.094	.307*								
20) DT Difficulty	.177	-.142	.030	.049	.261*	.426*							
21) SP Efficacy Belief	-.231$^+$	.024	-.020	.091	-.239$^+$	-.088	-.019						
22) FA Efficacy Belief	.020	-.002	.108	.099	.071	-.378*	-.086	.336*					
23) DT Efficacy Belief	-.113	.018	-.015	.109	-.029	-.126	-.299*	.455*	.566*				
24) MIA Achievement	.058	.058	-.023	.078	.213$^+$	-.041	.218$^+$	.187	.304*	.264*			
25) MIA Locus	-.183	-.149	-.219$^+$	-.068	.220$^+$	-.005	.159	-.032	.138	.184	.585*		
26) MIA Internal Strategies	-.121	.034	-.109	-.058	-.030	-.135	.117	.179	.029	.004	.392*	.412*	
27) MIA External Strategies	.067	.081	.103	.008	-.087	.037	.179	.259*	.169	.092	.079	.106	.433*

Note. Spearman *rho* used for self-reported cluster use, difficulty and efficacy belief reports which were made on ordinal Likert scales. Pearson *r* is used for all other variables. SP = spontaneous list; FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; MIA = Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983).

* $p \leq .05$; $^+$ $.05 < p \leq .10$.

Table C3

Full Correlation Matrix for Older Adults in the Individual-Words Format

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) SP RR													
2) FA RR	.164												
3) DT RR	.035	.279*											
4) SP LBC	.893*	.250+	.129										
5) FA LBC	.245+	.722*	.207	.357*									
6) DT LBC	.208	.266*	.816*	.236+	.279*								
7) SP Self-Reported Cluster Use	.445*	.401*	.071	.580*	.414*	.142							
8) FA Self-Reported Cluster Use	.133	.476*	.257+	.209	.627*	.306*	.564*						
9) DT Self-Reported Cluster Use	.086	.295*	.424*	.094	.305*	.495*	.237+	.490*					
10) SP Recall	.491*	.387*	.206	.682*	.472*	.169	.626*	.300*	.159				
11) FA Recall	.245+	.419*	.116	.333*	.851*	.195	.336*	.503*	.244+	.355*			
12) DT Recall	.173	.339*	.612*	.201	.274*	.829*	.119	.295*	.590*	.174	.245+		
13) Education Years	.010	-.031	.248+	-.012	.163	.213	.104	.170	.343*	-.040	.238+	.187	
14) Vocabulary	.245+	.030	.142	.248+	.238+	.207	.256*	.398*	.244+	.225+	.259*	.186	.452*
15) Speed	-.032	.245+	.066	.054	.267*	.171	.024	.174	.323*	.217+	.329*	.389*	.043
16) WMC	-.173	.375*	.201	-.051	.414*	.165	.133	.399*	.379*	.193	.249+	.161	-.009
17) RT Clustering Costs	.006	.303*	.102	.151	.192	.086	.285*	.183	.240+	.281*	.025	.189	.010
18) SP Difficulty	-.199	-.040	-.024	-.292*	-.040	-.020	-.222+	-.137	.042	-.226+	.029	.089	-.033
19) FA Difficulty	-.114	-.211	-.152	-.072	-.404*	-.233+	-.003	-.320*	-.125	.075	-.373*	-.173	-.073
20) DT Difficulty	-.020	.052	-.189	-.008	.053	-.286*	.139	-.082	-.264*	.123	.111	-.442*	-.019
21) SP Efficacy Belief	.013	.159	.253+	.125	.011	.154	.356*	.198	.215+	.151	-.116	.102	.151
22) FA Efficacy Belief	.099	.218+	.301*	.131	.303*	.292*	.396*	.411*	.377*	.058	.289*	.262*	.234+
23) DT Efficacy Belief	.145	.101	.253+	.139	.122	.334*	.121	.249+	.456*	-.042	.095	.441*	.229+
24) MIA Achievement	-.021	.018	-.117	-.016	-.049	-.199	.143	.098	.123	.007	-.105	-.127	-.047
25) MIA Locus	-.055	.011	-.290*	-.028	-.078	-.386*	-.030	-.208	-.205	-.023	-.251+	-.342*	-.218+
26) MIA Internal Strategies	.134	-.048	.088	.110	-.043	-.010	.241+	-.002	.153	.165	-.088	.003	.199
27) MIA External Strategies	.147	.113	.116	.133	.151	.067	.293*	.187	.224+	.096	.161	.185	.231+

Table C3 (continued)

Measure	14	15	16	17	18	19	20	21	22	23	24	25	26
1) SP RR													
2) FA RR													
3) DT RR													
4) SP LBC													
5) FA LBC													
6) DT LBC													
7) SP Self-Reported Cluster Use													
8) FA Self-Reported Cluster Use													
9) DT Self-Reported Cluster Use													
10) SP Recall													
11) FA Recall													
12) DT Recall													
13) Education Years													
14) Vocabulary													
15) Speed	-.138												
16) WMC	-.062	.330*											
17) RT Clustering Costs	-.007	.084	.254^{+}										
18) SP Difficulty	-.202	.176	.055	-.030									
19) FA Difficulty	-.154	-.054	-.129	.229^{+}	.183								
20) DT Difficulty	-.060	-.012	-.072	.060	.071	.437*							
21) SP Efficacy Belief	-.036	-.144	.049	.166	-.159	.085	-.104						
22) FA Efficacy Belief	.105	.117	.210	.229^{+}	-.037	-.160	.008	.636*					
23) DT Efficacy Belief	.012	.220^{+}	.119	.128	.075	.006	-.216^{+}	.470*	.719*				
24) MIA Achievement	-.160	.156	.116	.004	-.317*	-.079	-.092	.245^{+}	.290*	.172			
25) MIA Locus	-.275*	.036	.064	-.017	-.188	.168	.190	.078	-.099	-.064	.474*		
26) MIA Internal Strategies	.047	.198	.046	-.088	-.034	.029	.173	.284*	.302*	.210	.466*	.302*	
27) MIA External Strategies	.351*	.152	.163	.192	.053	-.101	.029	.249^{+}	.501*	.420*	.155	-.102	.487*

Note. Spearman *rho* used for self-reported cluster use, difficulty and efficacy belief reports which were made on ordinal Likert scales. Pearson *r* is used for all other variables. SP = spontaneous list; FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; MIA = Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983).

* $p \leq .05$; $^{+}$ $.05 < p \leq .10$.

Table C4

Full Correlation Matrix for Older Adults in the Whole-List Format

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) SP RR													
2) FA RR	.236^{+}												
3) DT RR	-.055	.018											
4) SP LBC	.828*	.247^{+}	.077										
5) FA LBC	.282*	.461*	.096	.340*									
6) DT LBC	-.040	.078	.429*	.022	.453*								
7) SP Self-Reported Cluster Use	.631*	.103	.110	.714*	.289*	-.016							
8) FA Self-Reported Cluster Use	.086	-.152	.013	.195	.348*	.064	.407*						
9) DT Self-Reported Cluster Use	-.023	.008	-.160	.074	.261*	.273*	.210	.731*					
10) SP Recall	.484*	.180	.044	.713*	.452*	.264*	.674*	.440*	.380*				
11) FA Recall	.147	-.043	.069	.209	.868*	.438*	.262*	.509*	.338*	.406*			
12) DT Recall	-.062	.046	-.107	-.085	.421*	.806*	.002	.170	.470*	.180	.429*		
13) Education Years	-.060	-.189	.036	.116	.112	.087	.125	.488*	.411*	.248^{+}	.248^{+}	.072	
14) Vocabulary	-.048	-.057	-.012	.046	.264*	.129	.039	.417*	.296*	.217^{+}	.369*	.105	.621*
15) Speed	-.095	.041	-.263^{+}	.136	.293*	.139	.193	.471*	.573*	.373*	.349*	.280*	.241^{+}
16) WMC	-.121	.071	-.249^{+}	.000	.385*	.233^{+}	-.034	.386*	.456*	.225^{+}	.426*	.326*	.244^{+}
17) RT Clustering Costs	.111	.130	.171	.166	.192	.352*	.000	.065	.128	.174	.124	.268*	-.011
18) SP Difficulty	-.315*	-.175	-.209	-.451*	-.274*	-.190	-.535*	-.354*	-.386*	-.597*	-.216^{+}	-.144	-.271*
19) FA Difficulty	-.105	.047	-.074	-.177	-.252^{+}	-.055	-.199	-.291*	-.252^{+}	-.235^{+}	-.313*	-.055	-.257*
20) DT Difficulty	.215^{+}	.124	.017	.200	.219^{+}	-.363*	.188	.029	-.302*	.121	.135	-.385*	-.074
21) SP Efficacy Belief	.224^{+}	.217^{+}	.002	.217^{+}	.195	-.201	.394*	.234^{+}	.164	.184	.105	-.092	.012
22) FA Efficacy Belief	.208	.053	.100	.272*	.421*	.222^{+}	.358*	.499*	.356*	.293*	.428*	.261*	.170
23) DT Efficacy Belief	.080	.156	.026	.145	.364*	.116	.193	.375*	.374*	.188	.318*	.168	.082
24) MIA Achievement	-.049	.031	.095	.016	.138	.071	.085	.327*	.299*	-.001	.121	.026	.026
25) MIA Locus	.141	-.019	.093	.044	.168	.121	.194	.112	.014	.150	.201	.141	-.050
26) MIA Internal Strategies	.036	.027	.103	.133	.241^{+}	.091	.146	.351*	.292*	.061	.213	.068	.088
27) MIA External Strategies	-.141	.003	.022	-.125	.139	.065	-.134	.294*	.213	-.137	.165	.110	.113

Table C4 (continued)

Measure	14	15	16	17	18	19	20	21	22	23	24	25	26
1) SP RR													
2) FA RR													
3) DT RR													
4) SP LBC													
5) FA LBC													
6) DT LBC													
7) SP Self-Reported Cluster Use													
8) FA Self-Reported Cluster Use													
9) DT Self-Reported Cluster Use													
10) SP Recall													
11) FA Recall													
12) DT Recall													
13) Education Years													
14) Vocabulary													
15) Speed	.318*												
16) WMC	.306*	.613*											
17) RT Clustering Costs	-.077	.139	.198										
18) SP Difficulty	-.217+	-.316*	-.156	-.029									
19) FA Difficulty	-.133	-.186	-.118	-.092	.484*								
20) DT Difficulty	.118	.070	.091	-.128	.205	.234+							
21) SP Efficacy Belief	-.037	.314*	.223+	.175	-.280*	-.160	.106						
22) FA Efficacy Belief	.107	.432*	.349*	.359*	-.273*	-.266*	.059	.583*					
23) DT Efficacy Belief	.018	.326*	.353*	.142	-.253+	-.108	-.051	.371*	.717*				
24) MIA Achievement	.029	.079	.209	.012	.089	-.110	-.199	.203	.322*	.426*			
25) MIA Locus	-.146	.003	.171	.070	-.153	-.035	-.137	.090	.230+	.288*	.394*		
26) MIA Internal Strategies	.069	.119	.242+	.027	-.055	-.142	.044	.221+	.424*	.332*	.465*	.004	
27) MIA External Strategies	.296*	.123	.169	.043	.093	-.204	-.062	.008	.298*	.240+	.246+	.022	.486*

Note. Spearman *rho* used for self-reported cluster use, difficulty and efficacy belief reports which were made on ordinal Likert scales. Pearson *r* is used for all other variables. SP = spontaneous list; FA = full-attention clustering-instructed list; DT = dual-task clustering-instructed list; MIA = Metamemory in Adulthood Questionnaire (Dixon & Hultsch, 1983).

* $p \leq .05$; $^{+}$ $.05 < p \leq .10$.

APPENDIX D

COMPARISON OF ARC AND RR CLUSTERING INDEX

Table D1

Means, Standard Errors, Computability, and Distributional Properties of ARC and RR Scores

	Younger Adults		Older Adults	
List	Individual Words	Whole List	Individual Words	Whole List
		Spontaneous List		
ARC	.27 (.06) n = 66 Skew = -1.59 Kurtosis = 6.93	.19 (.06) n = 64 Skew = 0.15 Kurtosis = -0.72	.36 (.07) n = 58 Skew = -1.91 Kurtosis = 6.69	.44 (.06) n = 54 Skew = -0.50 Kurtosis = -0.16
RR	.37 (.02) n = 66 Skew = 0.23 Kurtosis = -0.62	.31 (.03) n = 66 Skew = 0.50 Kurtosis = -0.91	.38 (.03) n = 60 Skew = -0.20 Kurtosis = -0.68	.42 (.03) n = 60 Skew = 0.16 Kurtosis = -0.70
		FA Clustering-Instructed List		
ARC	.70 (.04) n = 66 Skew = -1.45 Kurtosis = 1.97	.68 (.05) n = 66 Skew = -1.52 Kurtosis = 1.53	.72 (.06) n = 58 Skew = -4.33 Kurtosis = 24.88	.85 (.02) n = 58 Skew = -1.30 Kurtosis = 1.68
RR	.60 (.02) n = 66 Skew = -1.26 Kurtosis = 1.24	.61 (.03) n = 66 Skew = -1.27 Kurtosis = 0.74	.59 (.02) n = 60 Skew = -1.27 Kurtosis =2.50	.70 (.01) n = 59 Skew = -0.65 Kurtosis = 0.42
		DT Clustering-Instructed List		
ARC	.60 (.05) n = 62 Skew = -1.38 Kurtosis = 1.59	.63 (.05) n = 63 Skew = -1.07 Kurtosis = 0.29	.49 (.10) n = 46 Skew = -2.17 Kurtosis = 5.54	.71 (.06) n = 47 Skew = -1.69 Kurtosis = 2.32
RR	.52 (.02) n = 66 Skew = -0.56 Kurtosis = 0.04	.54 (.03) n = 66 Skew = -0.92 Kurtosis = -0.10	.43 (.03) n = 53 Skew = -0.56 Kurtosis =-0.60	.65 (.03) n = 56 Skew = -0.61 Kurtosis = 0.95

Note. ARC = Adjusted ratio of clustering (Roenker et al., 1971); RR = ratio of repetition (Bousfield, 1953).

It is evident from the comparison in Table D1 that the RR score is favorable in terms of score computability and distributional properties. The ARC score distributions show great kurtosis which is primarily due to outlying negative scores which are not bounded (maximum score is +1 but negative scores can be below -1; see Frankel & Cole, 1971, for a criticism of this property of the ARC score). RR scores have a lower bound of 0 and an upper bound of 1, resulting in more normal distributions. Skewed distributions with large kurtosis reduce statistical power in ANOVA (Levine & Dunlap, 1982) or require the use of less powerful nonparametric tests. This loss of power is particularly concerning combined with the reduced sample size for ARC analysis. Missing scores cannot be assumed to be missing at random (see Footnote 5), requiring listwise deletion which results in exclusion of 42 participants (28%) for ARC analysis but only 12 (5%) for RR analysis.

ARC and RR scores cannot be directly compared because ARC indicates the proportion clustered between chance and maximum possible clustering whereas RR simply indicates the proportion of the output that is clustered. Importantly, the pattern across cells in terms of higher and lower scores is consistent across the two measures and an exploratory ANOVA analysis of ARC scores let to the same conclusions as the presented ANOVA of RR scores. Nonparametric Spearman *rho* correlations indicated high consistency of the two measures, $r(242) = .952$ for the spontaneous list, $r(248) = .866$ for the FA clustering-instructed list, and $r(218) = .834$ for the DT clustering-instructed list, all $p < .001$.

APPENDIX E

ADDITIONAL RECALL ANALYSES

Additional analyses of the recall outputs as well as of the free recall of category labels and the category-cued recall provide information about semantic organization of the words in memory. Table E1 displays means and standard errors for these measures, which were all analyzed with the same general mixed ANOVA model as the primary data.

Table E1

Means and Standard Errors for Various Aspects of Recall Performance by List, Age Group, and Presentation Format

	Younger Adults		Older Adults	
List	Individual Words	Whole List	Individual Words	Whole List
	Proportion of Free Recall Intrusions			
Spontaneous	.04 (.01)	.05 (.01)	.07 (.02)	.06 (.02)
FA Clustering Instructed	.04 (.01)	.03 (.01)	.05 (.01)	.05 (.01)
DT Clustering Instructed	.06 (.01)	.07 (.02)	.14 (.02)	.12 (.02)
	Number of Categories Represented in Recall Output			
Spontaneous	3.91 (.04)	3.85 (.04)	3.72 (.07)	3.53 (.09)
FA Clustering Instructed	3.95 (.03)	3.91 (.04)	3.52 (.09)	3.32 (.10)
DT Clustering Instructed	3.35 (.11)	3.49 (.09)	2.37 (.14)	2.40 (.13)
	Number of Categories Freely Recalled			
Spontaneous	3.16 (.11)	3.11 (.16)	2.81 (.15)	2.46 (.17)
FA Clustering Instructed	3.79 (.05)	3.60 (.10)	3.22 (.12)	3.11 (.12)
DT Clustering Instructed	3.16 (.12)	3.13 (.14)	2.29 (.15)	2.41 (.14)
	Proportion of Words per Recalled Category			
Spontaneous	.62 (.02)	.57 (.02)	.53 (.02)	.54 (.03)
FA Clustering Instructed	.69 (.02)	.74 (.02)	.58 (.02)	.68 (.02)
DT Clustering Instructed	.54 (.02)	.62 (.02)	.41 (.03)	.56 (.02)
	Difference Proportion Category-Cued – Freely Recalled			
Spontaneous	-.03 (.01)	-.04 (.01)	.01 (.02)	-.03 (.02)
FA Clustering Instructed	-.01 (.01)	-.02 (.01)	.00 (.01)	.01 (.02)
DT Clustering Instructed	.04 (.01)	.01 (.01)	.07 (.02)	.04 (.02)

Note. FA = full attention; DT = dual task.

Intrusions

The proportion of intrusions (i.e., words not from the list in the recall output) significantly varied across lists, $F(1.79, 443.07) = 21.47$, $MSE = .01$, $\eta_p^2 = .08$, $p < .001$. Compared to the spontaneous list, participants had fewer intrusions on the FA clustering-instructed list, $t(251) = 2.34$, $d = 0.15$, $p = .020$, but more on the DT clustering-instructed list, $t(251) = 4.04$, $d = 0.26$, $p < .001$. That is, the proportion of intrusions mirrored list differences in recall performance with more intrusions when memory was poor. There was a main effect of age group, $F(1, 248) = 11.23$, $MSE = .02$, $\eta_p^2 = .04$, $p = .001$, that interacted with list, $F(1.79, 443.07) = 4.23$, $MSE = .01$, $\eta_p^2 = .02$, $p = .019$. Numerically, older adults had more intrusions than younger adults on all lists but this age-related difference was small and not significant on the spontaneous list, $t(250) = 1.50$, $d = 0.19$, $p = .136$, and only marginal on the FA clustering-instructed list, $t(250) = 1.74$, $d = 0.22$, $p = .082$, but moderately sized and significant on the DT clustering-instructed list, $t(250) = 3.46$, $d = 0.44$, $p = .001$. That is, under divided attention older adults were particularly susceptible to intrusions. No other effects were significant, all $F<1$. That is, the proportion of intrusions did not differ with presentation format. Hence, the better recall in the whole-list conditions on the clustering-instructed lists was present at a similar level of intrusions. Notably, the majority of intrusions were semantically related to the study words, occasionally even the output of a category label (e.g., "flower").

Memory for the Semantic Categories

Memory for the semantic categories of list words (maximum 4) was assessed in two ways: (1) By examining the number of categories represented by at least one word in the recall output, and, (2) by examining participants' free recall of category labels after the recall test for each list. The first measure indicates whether participants narrowed down on just a few categories or recalled more broadly from the entire range of categories in the study list whereas the second measure gives a direct indication in as much the categorical structure of a list was abstracted (independent of whether words from each category were recalled). Regarding the number of categories represented in the recall output, the main effect of list was significant, $F(1.73, 429.50) = 135.05$, $MSE = .48$, $\eta_p^2 = .35$, $p < .001$. The average number of categories represented in the recall output did not differ between the spontaneous and the FA clustering-instructed list, $t(251) = 1.57$, $p = .117$. Notably, means were above 3.5 on the spontaneous list in all conditions, that is most participants spontaneously recalled words from all four categories and this ceiling performance hence may have precluded additional benefits of clustering instructions in this measure. Divided attention significantly reduced the number of categories represented in the recall output compared to the spontaneous list, $t(251) = 12.08$, $d = 0.82$, $p < .001$. There was a main effect of age group, $F(1, 248) = 112.32$, $MSE = .61$, $\eta_p^2 = .31$, $p < .001$, that interacted with list $F(1.73, 429.50) = 23.96$, $MSE = .48$, $\eta_p^2 = .09$, $p < .001$. Older adults' recalled words from fewer categories than younger adults on all three lists, all $p < .001$, but this age-related difference was smaller on the spontaneous list ($d = 0.50$) than when clustering was instructed (FA $d = 0.93$; DT $d = 1.11$). No other effects were significant, all $F \leq 2.18$, $p \geq .122$. That is, there were no effects involving the presentation format manipulation.

Free category recall also varied with list, $F(1.82, 450.44) = 36.27$, $MSE = .87$, $\eta_p^2 = .14$, $p < .001$. Compared to the spontaneous list, participants recalled more categories on the FA clustering-instructed list, $t(251) = 7.08$, $d = 0.46$, $p < .001$. Even though these two lists had not differed in the number of categories represented in the recall output, this measure reveals that participants better represented the list categories after clustering instructions. There was no difference between free category recall on the spontaneous and the DT clustering-instructed list, $t(251) = -1.58$, $p = .116$. That is, even though the number of categories represented in the recall output had been reduced on the DT compared to the spontaneous list, participants were able to freely name as many categories as on the spontaneous list, suggesting that after clustering instructions participant had paid attention to the semantic properties of the list under DT. Older adults recalled fewer categories, $F(1, 248) = 38.96$, $MSE = 1.78$, $\eta_p^2 = .14$, $p < .001$, across the three lists. No other effects were significant, all $F \leq 2.12$, $p \geq .126$. That is, there were again no effects involving presentation format. Generally, category recall was fairly high, suggesting that participants successfully abstracted the semantic list properties.

Proportion of Items Recalled Per Category

For each participant and each list, the average proportion of words per each category of which at least one word was recalled was computed. The more words were semantically grouped, the more completely should list categories be recalled. There was a significant main effect of list, $F(1.47, 459.48) = 72.89$, $MSE = .02$, $\eta_p^2 = .23$, $p < .001$. Compared to the spontaneous list, categories were recalled more completely on the FA clustering-instructed list, $t(251) = 9.51$, $d = 0.60$, $p < .001$, and less completely on the DT clustering-instructed list, $t(251) = 2.21$, $d = 0.16$, $p = .028$. Older adults recalled categories less completely overall, $F(1, 248) = 27.05$, $MSE = .04$, $\eta_p^2 = .10$, $p < .001$, and this age effect was invariant across the lists, $F(1.85,459.47) = 1.35$, $p = .259$. There was a main effect of presentation format, $F(1, 248) = 14.84$, $MSE = .04$, $\eta_p^2 = .06$, $p < .001$, that interacted with list $F(1.85, 459.48) = 17.22$, $MSE = .02$, $\eta_p^2 = .07$, $p < .001$. There was no difference in the average proportion of words per category recalled on the spontaneous list, $t(250) = 1.04$, $p = .299$. Once clustering was instructed, however, categories were recalled more completely in the whole-list conditions, FA $t(250) = 4.11$, $d = 0.52$, $p < .299$, and DT $t(250) = 4.96$, $d = 0.63$, $p < .001$. That is, more complete recall of categories appears to underlie the recall-benefit of the whole-list format on the clustering-instructed lists. This finding is also in line with the LBC analyses as LBC increases with more category repetitions and longer category cluster which both imply that more members of a category are recalled. Lastly, presentation format significantly interacted with age group, $F(1.85, 459.48) = 17.22$, $MSE = .02$, $\eta_p^2 = .07$, $p < .001$. Although numerically a higher proportion of words from a category was recalled in the whole-list format overall (i.e., averaged across the three lists) in both age groups, this overall difference was only reliable in the older adults, $t(118) = 4.03$, $d = 0.74$, $p < .001$, but not in the younger adults, $t(130) = 1.33$, $d = 0.23$, $p = .186$. The three-way interaction was not significant, $F<1$, that is the Presentation Format x List interaction held for both younger and older adults. That is, when instructed to cluster, all participants benefitted from the whole-list format in terms of achieving more complete recall of list categories with stronger benefits in older adults. Consequently, the overall age-related difference in the completeness of category recall was smaller in the whole-list format (d = 0.39) than in the individual-words format (d = 0.93).

Category-Cued Recall

After free recall of each study list, participants also completed a second recall cued with the list categories. Results from the analysis of the mean proportion of category-cued recall did not differ from that of the proportion freely recalled. Of particular interest is to what extent the category cues facilitated recall. Therefore, difference scores were computed (proportion cued recall – proportion free recall) with positive scores reflecting a benefit of cueing. Mean difference scores are presented in Table E1. When interpreting these scores it must be kept in mind that they followed the free recall of a given list with a self-paced delay for answering interim questions (see Appendix B). There was a main effect of list, $F(2, 496) = 30.03$, $MSE = .01$, $\eta_p^2 = .11$, $p < .001$. One-sample t-tests comparing the mean difference score against 0 revealed that there only was a benefit (difference score > 0) of category cues on the DT clustering-instructed list, $t(251) = 5.50$, $d = 0.35$, $p < .001$. Category cues did not benefit recall on the FA clustering-instructed list, t<1, and for the spontaneous list the cued proportion recalled was even slightly reduced compared to free recall, $t(251) = -3.19$, $d = 0.20$, $p = .002$. Perhaps category cues slightly hindered recall on the spontaneous list because participants did not primarily organize the list in a semantic manner. There was further a main effect of age group, $F(1, 248) = 6.76$, $MSE = .02$, $\eta_p^2 = .03$, $p = .010$, that did not interact with list. Older adults overall benefitted from category cues, one-sample $t(119) = 2.03$, $d = 0.19$, $p = 0.45$, whereas younger adults did not, one-sample $t(132) = -1.57$, $p =$

.119. The overall benefit in older adults amounted to only 2% though and as is evident from Table E1 was only present on the DT list, consistent with the main list effect that did not interact with age group, *F*<1. The main effect of presentation format was marginally significant, $F(1, 248) = 3.72$, $MSE = .02$, $\eta_p^2 = .02$, $p = .055$. Overall, there was only a benefit of category cues in the individual-words format ($M = .01$, $SE = .01$), $t(125) = 2.07$, $d = 0.18$, $p = .041$, but not in the whole-list format ($M = -.00$, $SE = .01$), $t<1$. No interactions involving presentation format were significant, all $F \leq 1.69$, $p \geq .186$. Again, the benefit in the individual-words format was present on the DT list only and very small (1% on average). Taken together, category cues did not strongly benefit recall of the study lists, presumably participants cued themselves sufficiently in the free recall.

APPENDIX F

STRATEGY USE ON THE SPONTANEOUS LIST

Use of semantic clustering was the primary focus of the present study but some information on other strategies participants used on the spontaneous list was additionally obtained. In particular, after study of the spontaneous list, participants were presented with a checklist of strategy options and asked to check any strategies they had used during study of the first spontaneous list (see Appendix B). Table F1 presents the proportion of participants in each condition that checked a strategy option. These responses reveal if there were strategic differences between the conditions not related to semantic clustering use. Further, they help determine if the older adult sample was typical in terms of strategic behavior or unusually strategic, which is in turn informative for understanding the lack of age-related differences in clustering use.

Table F1

Checklist Strategy Reports and Serial Clustering Index for Spontaneous List

	Younger Adults		Older Adults	
	Individual Words	Whole List	Individual Words	Whole List
% Imagery	36.4	33.3	26.7	25.0
% Sentence	22.7	12.1	15.0	5.0
% Repetition	72.7	80.3	45.0	26.7
% Semantic Clustering	66.6	57.6	68.3	68.3
% Other Clustering	15.2	15.2	8.3	8.3
% Other	13.6	18.2	11.7	8.3
Number of Strategies	2.27 (.14)	2.17 (.12)	1.75 (.12)	1.42 (.12)
Serial Clustering Index	0.32 (.12)	0.63 (.24)	0.00 (.10)	0.04 (.10)

Note. Standard errors in parentheses.

For each strategy option, the responses (0 = not checked, 1 = checked) were analyzed with binomial regressions using dummy coding for age group (0 = younger, 1 = older) and presentation format (0 = individual words, 1 = whole list) as well as including their interaction term. Semantic clustering was checked by most participants and did not differ across conditions (all $p \geq .283$). Descriptively, fewer older adults reported using deep encoding strategies like imagery and sentence but this difference was not reliable (both $p \geq .244$) nor were there effects involving presentation format (all $p \geq .113$). For the more shallow encoding strategy repetition, a reliable age difference occurred with fewer older adults reporting its use, $B = -1.24$, $SE = .38$, $p = .002$. There was no effect of presentation format, $p = .306$, but a significant age group x presentation format interaction, $B = -1.81$, $SE = .28$, $p = .030$. Examined separately by age group, there was no presentation format effect in the younger adults ($p = .306$) but in the older adults, $B = 0.81$, $SE = .39$, $p = .038$. Specifically, fewer older adults reported using repetition in the whole-list compared to the individual-words format. For clustering by criteria other than semantic category and for use of other strategies, which were rare, no significant effects were found (all $p \geq .244$). Since participants were allowed to check multiple strategies on the checklist, the mean number of strategies used by participants could be compared across conditions with a 2 (Age Group: younger vs. older) x 2 (Presentation Format: individual words vs. whole list) ANOVA. See Table F1 for descriptive statistics. There was a main effect of age group, $F(1, 248) = 24.65$, $MSE = 1.03$, $\eta_p^2 = .09$, $p < .001$, with older adults using fewer strategies on average. The effect of presentation format was marginal, $F(1, 248) = 2.94$, $MSE = 1.03$, $\eta_p^2 = .01$, $p = .088$, with a tendency for more varied strategy use in the individual-words conditions. The interaction was not reliable, $F<1$. Thereby, this analysis replicates previous work of less varied strategy use in older adults (e.g., Bailey, Dunlosky, & Hertzog, 2009; Kuhlmann & Touron, 2012), suggesting that the current older adult sample was not unusually strategic. Thus, younger adults' superior recall despite similar semantic clustering on the spontaneous list may in part stem from pairing semantic clustering with other strategies. Notably, the shallow strategy repetition may be quite effective if it involves cumulative rehearsal of semantic groups. The number of strategies was indeed positively correlated with recall on the spontaneous list, $r(132) = .276$, $p = .001$, for younger adults and $r(120) = .276$, $p = .002$, for older adults.

Just like category repetitions in the recall output reveal semantic organization of study material, adjacent recall of words that occurred in neighboring serial positions in the study list reveal serial/temporal organization of the study material. A serial clustering index was computed according to the formula provided in Stricker et al. (2002), which deducts chance expectation of serial clustering from the observed serial clustering in a recall output. Serial clustering scores were analyzed with a 2 (Age Group: younger vs. older) x 2 (Presentation Format: individual words vs. whole list) univariate ANOVA. Younger adults' spontaneous recall outputs had a higher serial organization than those of older adults, $F(1, 248) = 12.36$, $MSE = 1.54$, $\eta_p^2 = .05$, $p = .001$. The effect of presentation format and the interaction were not significant, $F<1$. That is, younger adults were more likely to serially recall the study material in either presentation format, although it is notable that serial clustering was almost twice as high in the whole-list format. Notably, older adults' serial clustering indices were 0 indicating that they generally did not organize the material by study position. Semantic clustering results in negative serial recall indices because words from one semantic category are interspersed throughout the study list. Thereby, the similar or even higher spontaneous semantic clustering use in older adults on the spontaneous list might in part be due to younger adults' strong preference to organize serially rather than by semantic criteria. Witte et al. (1993) found a similar preference for serial organization in younger, which diminished across repeated study-recall trials of a categorizable list.